NPS62Sa77111

NAVAL POSTGRADUATE SCHOOL

Monterey, California

INTRODUCTION TO ACOUSTIC SIGNAL PROCESSING

by

George L. Sackman
LT William O. Davis, USN
LT Paul H. Donaldson, USN

November 1977

Approved for public release; distribution unlimited.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

Requests for the document must be referred to the Naval Postgraduate School, Monterey, California 93940, Code 023A.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS62Sa77111	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): INTRODUCTION TO ACOUSTIC SIGNAL PROCESSING	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): George L. Sackman LT William O. Davis, USN LT Paul H. Donaldson, USN	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School Monterey, California 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE: November 1977
	13. NUMBER OF PAGES: 163
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Signal Processing	FFT	Narrowband Analysis
Spectrum Analysis	Beamformer	
Correlation	Ambiguity Diagram	
DELTIC	Acoustic Signals	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Acoustic Signal processing is a short course in electrical signal processing fundamentals and their applications in the field of underwater acoustics. It contains an introduction to Fourier transforms and their properties, sampling and quantization, filters and bandwidth requirements, random signals and noise, and an introduction to four types of processing equipment; the DELTIC, energy detectors, correlation detectors, and

beamformers. Course objectives are given in terms of specific questions which a person completing the course should be able to answer. The course is designed to be presented to the personnel involved with the development, operation and employment of acoustic sensors to provide them with a better understanding of the operations accomplished by their equipment and to develop in them a better appreciation of the problems and limitations associated with signal detection in the underwater environment.

ABSTRACT

Acoustic Signal Processing is a short course in electrical signal processing fundamentals and their applications in the field of underwater acoustics. It contains an introduction to Fourier transforms and their properties, sampling and quantization, filters and bandwidth requirements, random signals and noise, and an introduction to four types of processing equipment; the DELTIC, energy detectors, correlation detectors, and beamformers. Course objectives are given in terms of specific questions which a person completing the course should be able to answer. The course is designed to be presented to the personnel involved with the development, operation and employment of acoustic sensors to provide them with a better understanding of the operations accomplished by their equipment and to develop in them a better appreciation of the problems and limitations associated with signal detection in the underwater environment.

TABLE OF CONTENTS

INTRODUCTION

These notes on Acoustic Signal Processing were originally developed as an introductory course intended for the Naval officer who has had the traditional training in sonar hardware. However, it is applicable to anyone involved in underwater acoustics --officer, enlisted, or civilian. An engineering background is not necessary; however, it is presumed that the concepts of calculus are somewhat familiar.

The course was designed to provide an overview of the principles involved in the application of Fourier analysis and statistics to acoustic signal processing and the operation of spectrum analyzers, energy detectors, correlation detectors, and beam formers. These processing methods comprise the core around which the acoustic sensor systems are built. In proceeding toward this goal, the course follows a logical approach of building from the basics of Fourier Transforms and Fourier Transform Properties. Following these two sections is a section on Sampling and Quantization and a section on Filters and Linear Systems. These latter two sections may be interchanged but both are required for the section on Random Signals, Power Spectral Density, and Noise. The methods of processing are presented last. They are DELTIC and FFT, Energy Detection, Correlation Detection, and Beam Forming. Where appropriate, specific hardware has been mentioned, but it is emphasized that the course is primarily devoted to principles and methods

that may be applied to various systems employed on either air, surface, or subsurface platforms.

COURSE OBJECTIVES

The overall objective of this course is to familiarize the student with the principles of acoustic signal processing as used in underwater sensor systems. The student who successfully completes the course will be knowledgeable to the degree indicated below.

The student should understand the theory of Fourier Analysis sufficiently well to do the following:

1. Explain the relationship of the Fourier Transform to the "time domain" and the "frequency domain".

2. Given a Fourier Transform pair, identify the operations expressed.

3. Given one long and one short rectangular pulse:

    a. Sketch the convolution of one with the other.

    b. Transform the given pulses from the time domain to the frequency domain, multiply them, and apply the inverse transform (approximately) to return to the time domain.

4. Given a square wave, sketch the autocorrelation and power spectrum.

The student should be sufficiently knowledgeable of signal processing fundamentals to do the following:

1. Show the difference between a digital and an analog signal.

2. Discuss how noise effects quantization of a signal.

3. State the Sampling Theorem and describe "aliasing".

4. Given a plot of signal and noise as power density versus frequency, show how filtering increases the ratio of signal power to noise power.

5. Relate filter bandpass, integration time, and frequency resolution.

6. Given a plot of transfer function versus frequency, sketch the impulse response and frequency response.

7. Define the statistical characteristics of a random signal and relate them to the signal voltage (or current) and power components.

8. List the assumptions necessary to apply statistical techniques to the processing of random signals.

9. Contrast the cross correlation of correlated signals with the cross correlation of uncorrelated signals. Show how the correlated signals are enhanced.

10. Explain the effects of filtering gaussian white noise. Discuss the relationship of these effects when processing signal and noise.

The student should be able to describe the DELTIC and the FFT, including:

1. The benefits of using a DELTIC in the processing of an acoustic signal and how a DELTIC works.

2. Explain the significance of Time-Bandwidth Product in terms of the integration time required and the frequency resolution of the output.

3. Compare the FFT with the Fourier Transform.

4. Compare a DELTIC with an FFT spectrum analyzer.

Describe energy detection by explaining:

1. How the system decides whether a target is present.

2. How the threshold setting effects the probability of detection, P(d), and the probability of false alarm, P(fa).

3. The significance of Receiver Operating Characteristics (ROC) curve.

The student should be able to explain how a correlator processes signals and how a correlator functions:

1. Describe the effect of doppler on correlation.

2. Sketch an ambiguity diagram and explain how it is derived.

3. Explain how doppler effects CW, FM, and PR pulses.

4. Describe how pulse type effects doppler resolution.

5. Describe how pulse type effects range resolution.

6. List the advantages of using an FM slide.

7. State the advantages and disadvantages of using pseudo-random pulses for search and tracking.

8. Given a two-hydrophone array being used with a correlation detector, sketch and explain how the signal direction is obtained, why a direction ambiguity exists, and how the ambiguity may be resolved.

9. Explain the limitations on the type of signals which may be processed passively by a correlation detector and why these limitations exist.

To explain how a beam former processes signals spacially, be able to:

1. Describe the method used by the beam former to obtain direction.

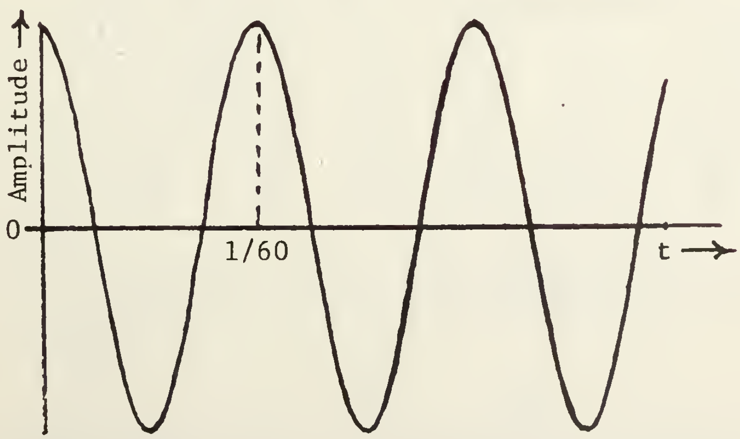
2. Show the relationship between aperture excitation, beam pattern, and spacial frequency

3. Given a hydrophone array in a beam former configuration, sketch the output of the beam former for a signal in the beam and a signal out of the beam. Show by comparing the outputs the meaning of array gain.

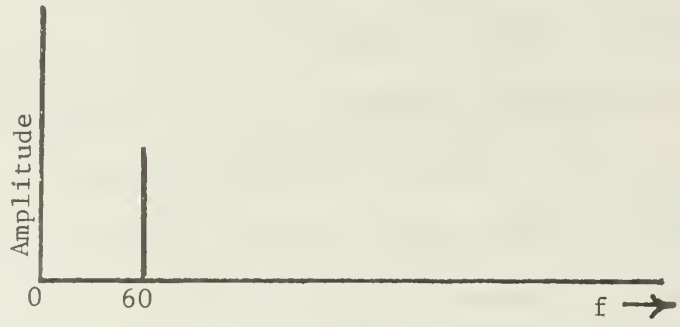
4. Describe the relationship of array gain to directivity index.

I. INTRODUCTION TO FOURIER TRANSFORMS

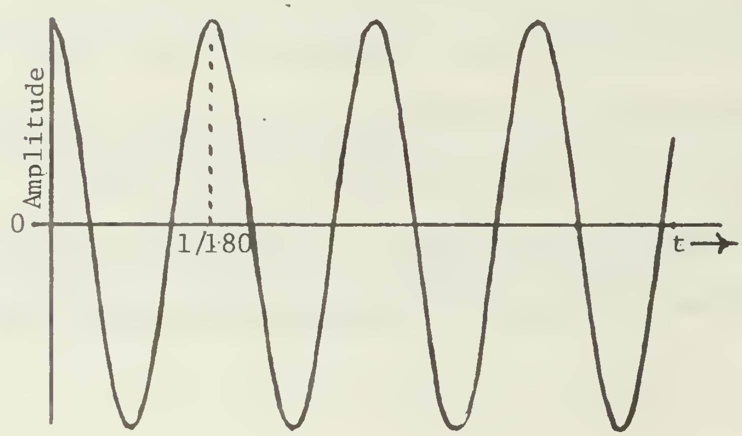
An introduction to signal processing requires, for understanding of many processing schemes, a working familiarity with Fourier transforms. To gain an understanding of what Fourier transforms are, what they can do, and how to use them, let us first examine a common type of signal. The electrical power used in the United States is 60 Hertz (cycles per second) alternating current. That is, the current changes its direction of flow in such a manner as to complete a cycle sixty times per second. If one plots the amplitude of the current in a given direction as a function of time, the result is a sinusoid. Since this is a description of the signal as a function of time, it is said to be the "time domain" description. If the signal is plotted as a function of frequency, it is called the "frequency domain" description, or "spectrum".

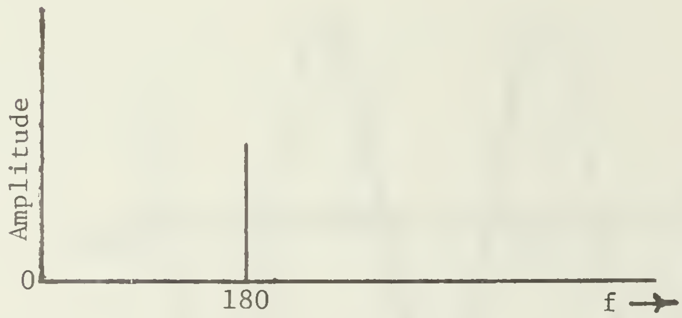


Because the signal consists of a single frequency, there is but one frequency represented on the f-domain plot.



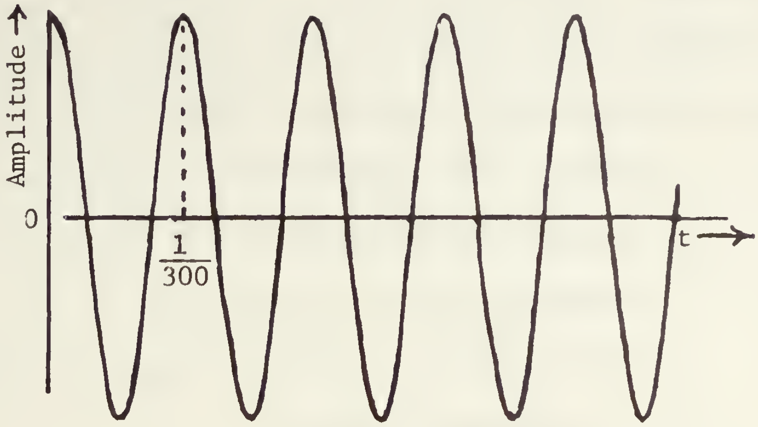
Now consider a different signal, an alternating signal of 180 Hz. The time domain and frequency domain plots are as shown:

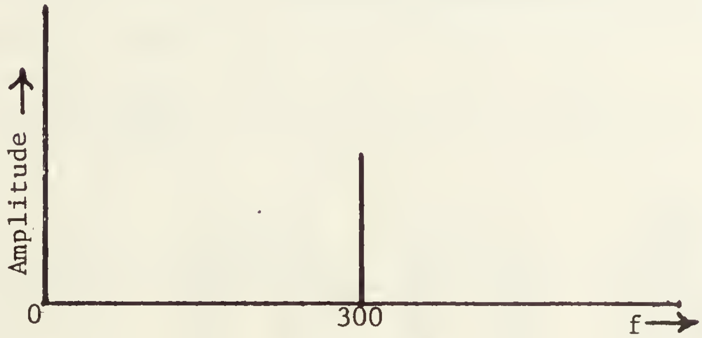


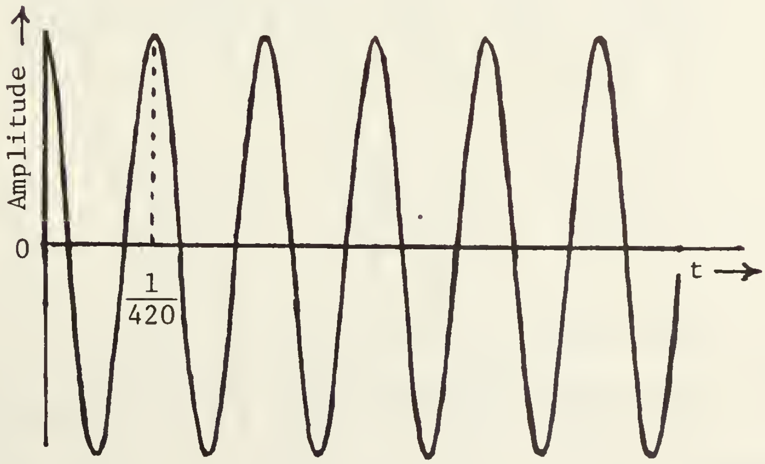


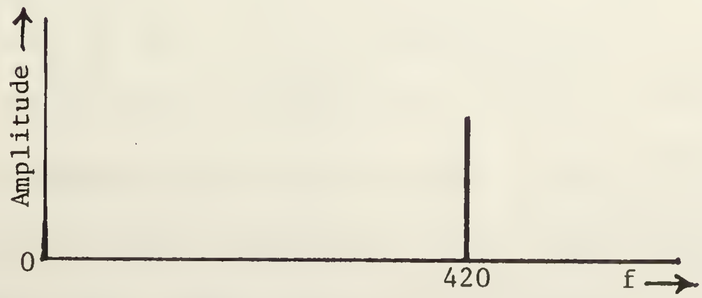
Note the way in which the changes in the signal are evidenced, both in the time domain and the frequency domain.

In the same way, two other signals, 300 Hz and 420 Hz are plotted:





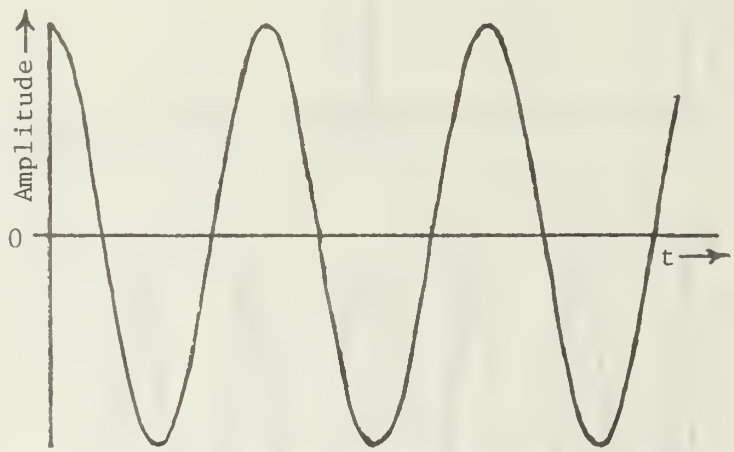




Thus far the signals shown have been of a single frequency. In general, there are also signals made up of many frequency components. Note at this point the concept of phase:

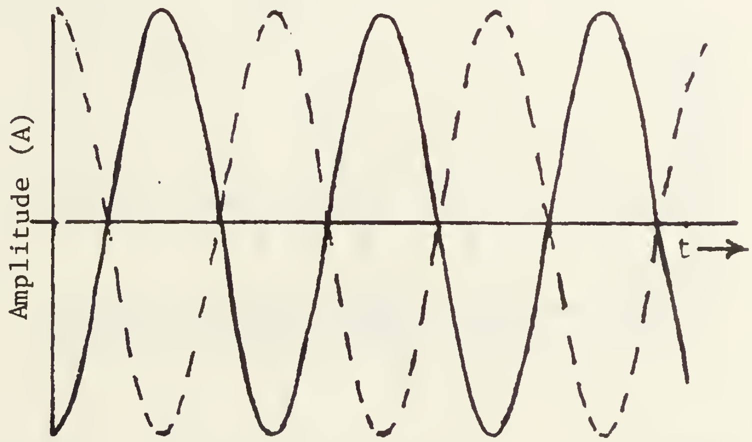
PHASE: The phase of a sinusoidal signal with respect to a reference signal is the relationship between corresponding parts of their cycles in the time domain.

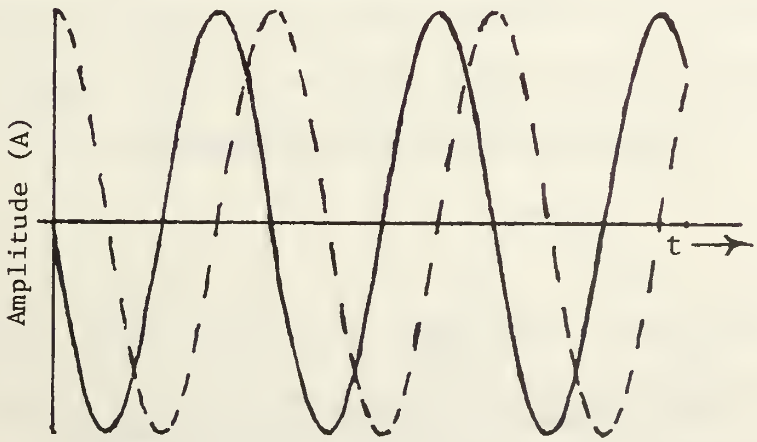
For example, a signal for reference is sketched below.



Comparing a signal with this reference, one can determine the phase of the new signal with respect to the reference. If the new signal crosses the t axis at the same time and in the same direction as the reference, it is said to be "in phase" with the reference, or to have a phase angle of 0°. If the new signal crosses the t axis on the down-swing when the reference signal crosses on the up-swing it is said to be "180° out of phase". The reason for using "180°" will be shown

later in this development. For the moment, however, it is only necessary to state that a signal goes through a phase angle of 360° each complete cycle. The phase of a signal can either lead or lag the reference, that is, the signal may cross the t-axis on the down-swing before or after the reference does. The first case would be "leading" and the second would be "lagging" with respect to the reference. Because of this, a signal is not generally described as having a phase

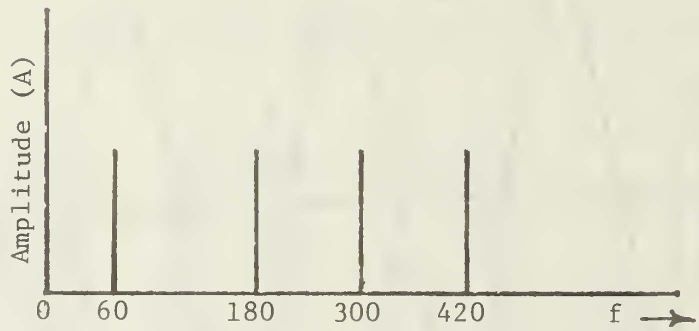




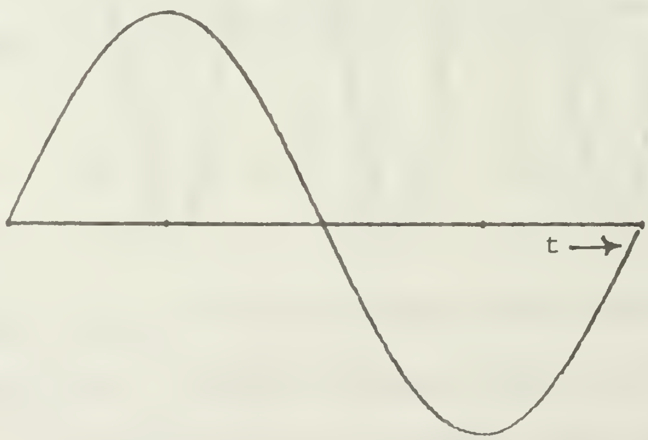
angle greater than 180°, since a 180° leading signal looks just the same as a 180° lagging signal. For example, a signal which is leading the reference by 270° looks the same as one

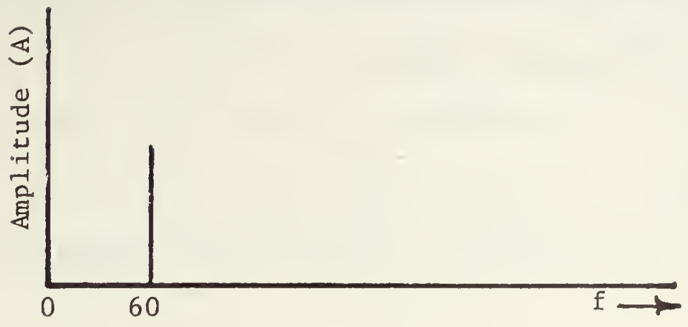
lagging the reference by 90°. In this case, it is usually described as the latter.

Returning to the signal with more than one frequency component, consider now the way in which this would be represented in the time and frequency domains. Take, for example, the four frequencies represented earlier: 60, 180, 300, and 420 Hz. If there were to be a signal composed of equal amplitude components of each of these frequencies, the frequency domain plot would look as shown below.

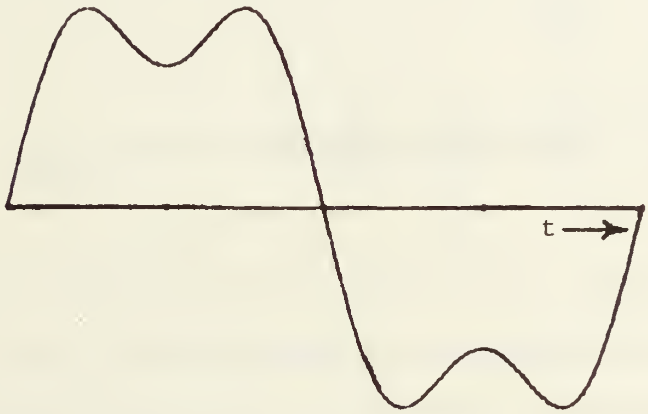


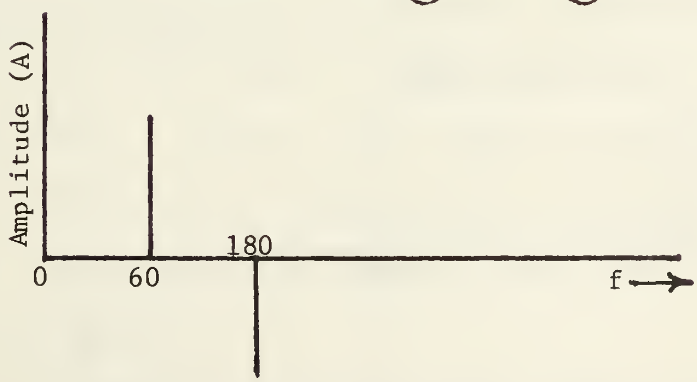
However, consider the time domain plot of this signal if, instead of equal amplitudes, judiciously chosen amplitude and phase components of the other frequencies are added to the 60 Hz frequency. First, plot the 60 Hz portion.





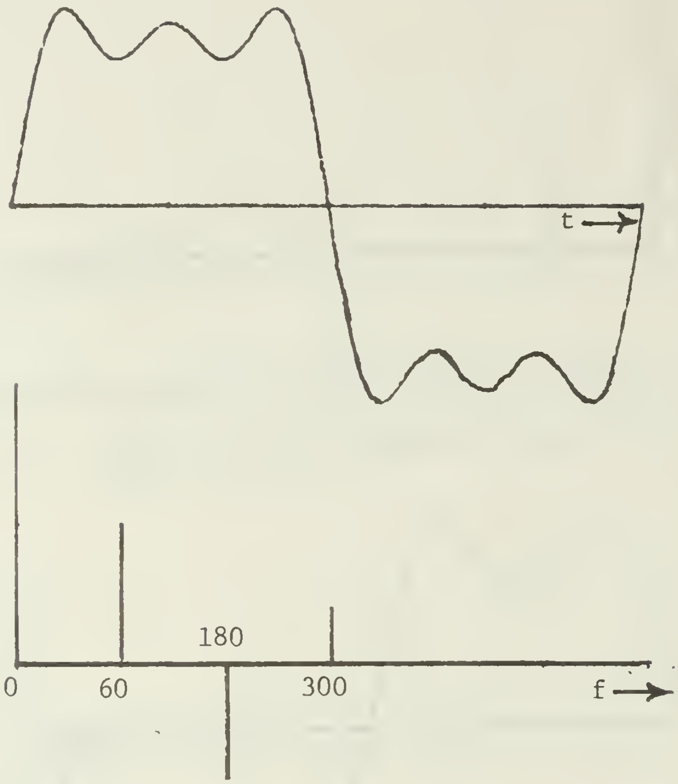
Next add the 180 Hz signal but with one-third the amplitude and shifted 180° out of phase at the peak.



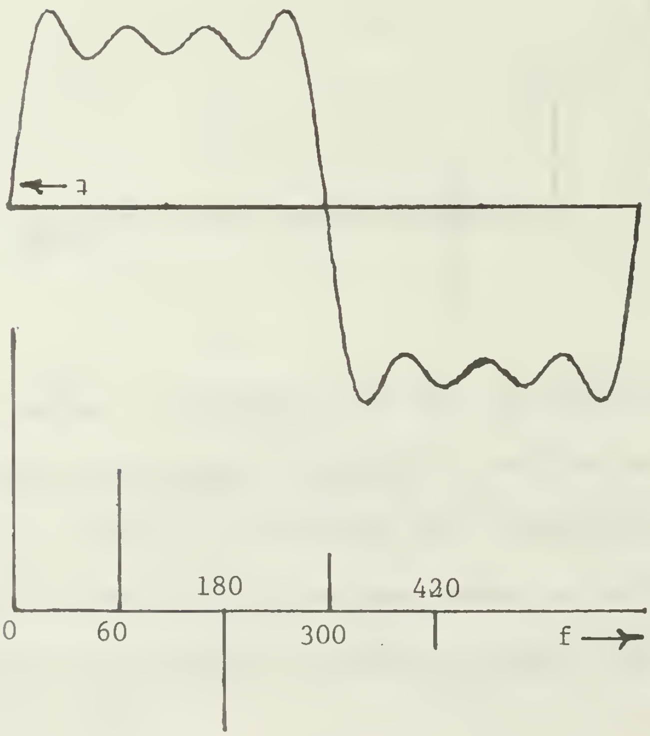


Note the fact that the 180 Hz component, added 180° out of phase, shows up on the frequency domain plot as a negative amplitude. The reason for this will become apparent. For now, the process is continued with the other frequencies. If one added to the signal a 300 Hz component, one-fifth the

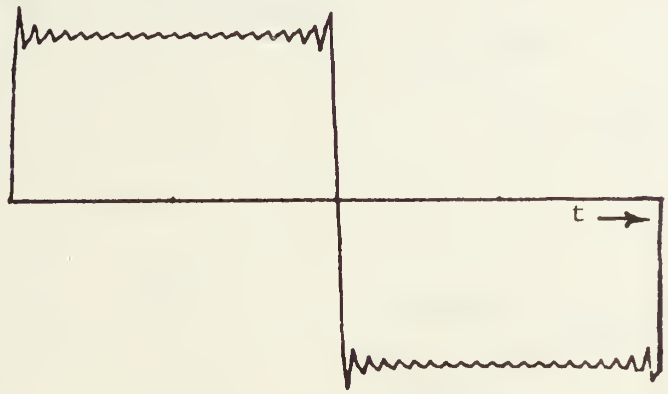
amplitude and in phase with the 60 Hz signal, it becomes:



Now adding one-seventh the amplitude with a 420 Hz component, 180° out of phase with the 60 Hz signal, it becomes:



thus proceeding to add carefully selected frequency components to the signal, the time domain signal becomes more and more like a square wave. With twenty components, the wave looks as illustrated. Therefore, one can see that by careful choice of component frequencies, and their respective amplitudes and

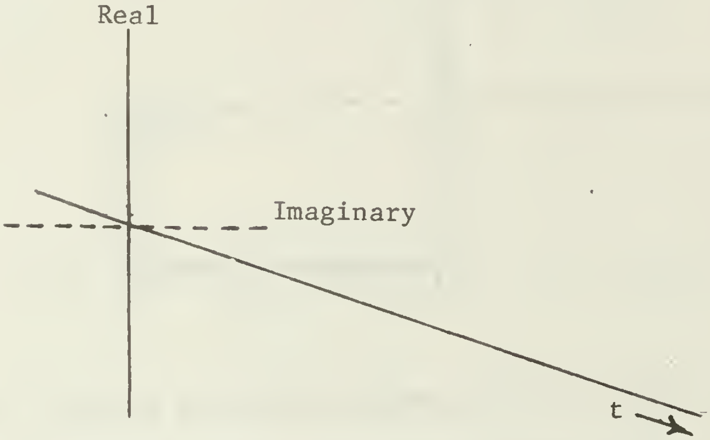


phases, a square wave can be generated from purely sinusoidal components. If an infinite number of components were used the resulting waveform would be a perfect square wave. This concept is the basis on which a large part of our signal processing theory rests. There is not time in this paper to demonstrate it, but it is a basic theorem of Fourier analysis that any shape waveform can be created from combinations of pure sinusoidal signals.

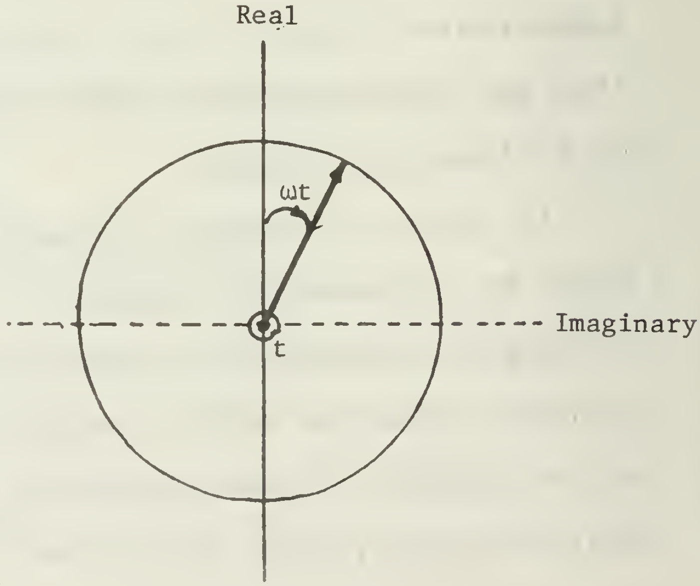
In order to simplify the mathematics involved later, it helps to introduce the concept of complex quantities. In solving the mathematical equations describing waves, one encounters solutions which contain terms multiplied by $\sqrt{-1}$. As can readily be seen, there is no real number which, when multiplied by itself, will give a number which is negative.

Thus, these quantities are known as "imaginary". In actuality, they are artifacts of the mathematics and are not directly related to tangible quantities. However, they prove to be of use in describing the behavior of wave phenomena.

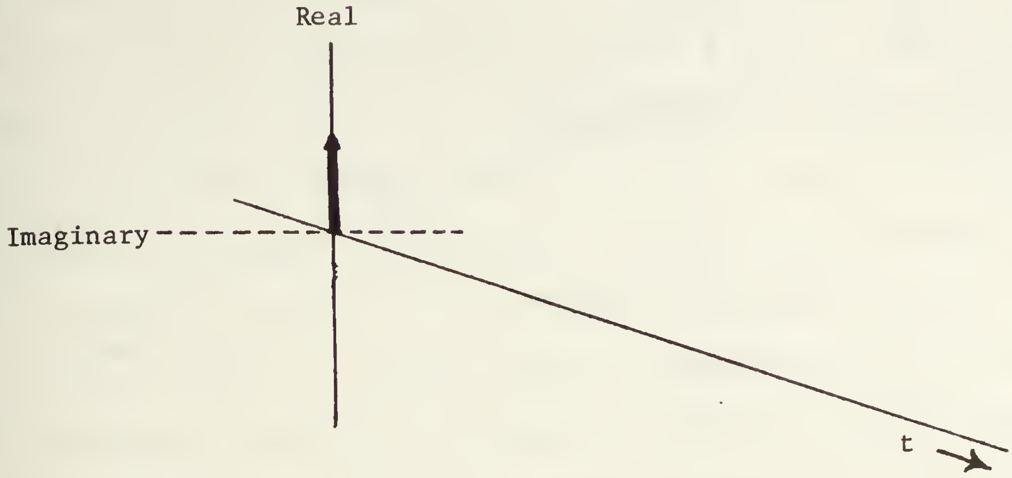
If "real" quantities are plotted on one axis, "imaginary" quantities on another, and time on a third axis, there appears a coordinate system as sketched:

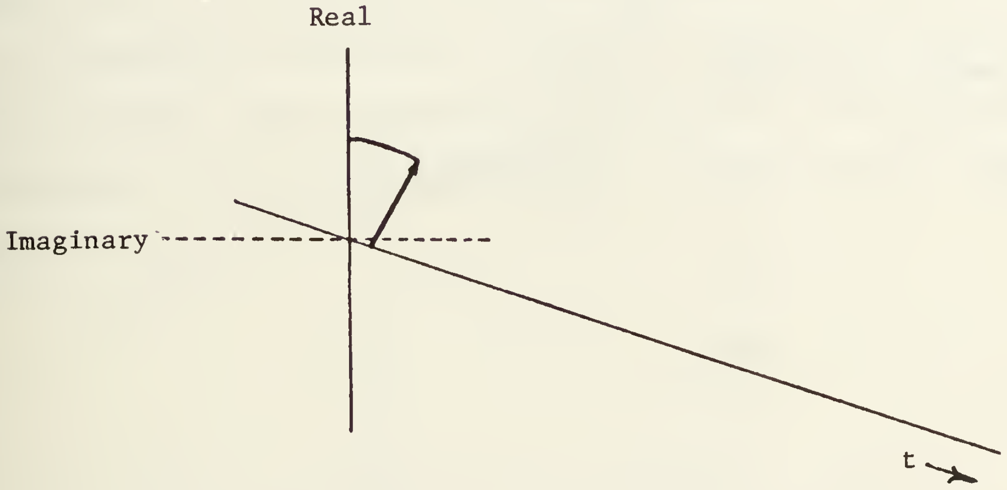


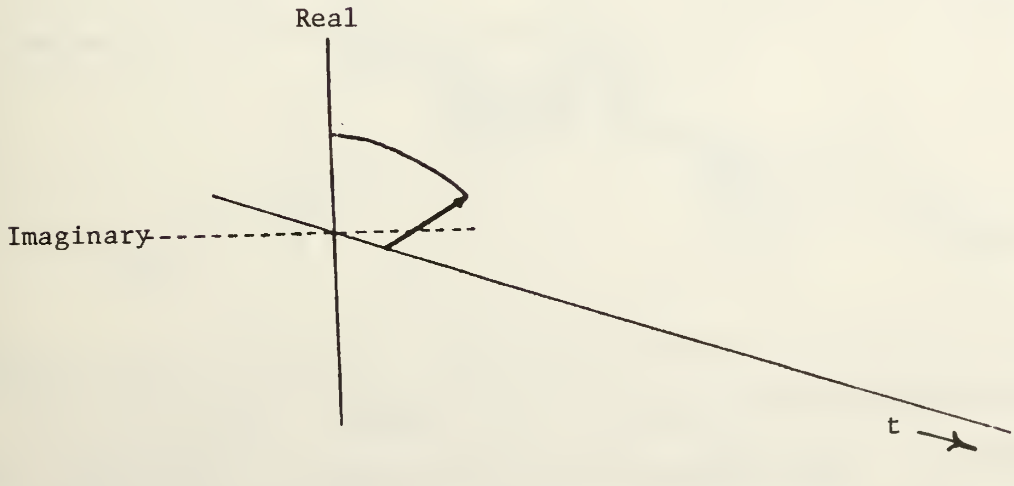
One of the results of the solutions described above is that a sinusoidal signal in the real plane (the plane formed by the real and time axes) can be generated by a unit vector rotating with a uniform angular velocity in the complex plane (formed by the real and imaginary axes). Looking at this vector down the time axis shows as illustrated in the diagram at the right. As this vector rotates, it is also moving down the time axis, so the path swept out by the tip of the vector describes a helical

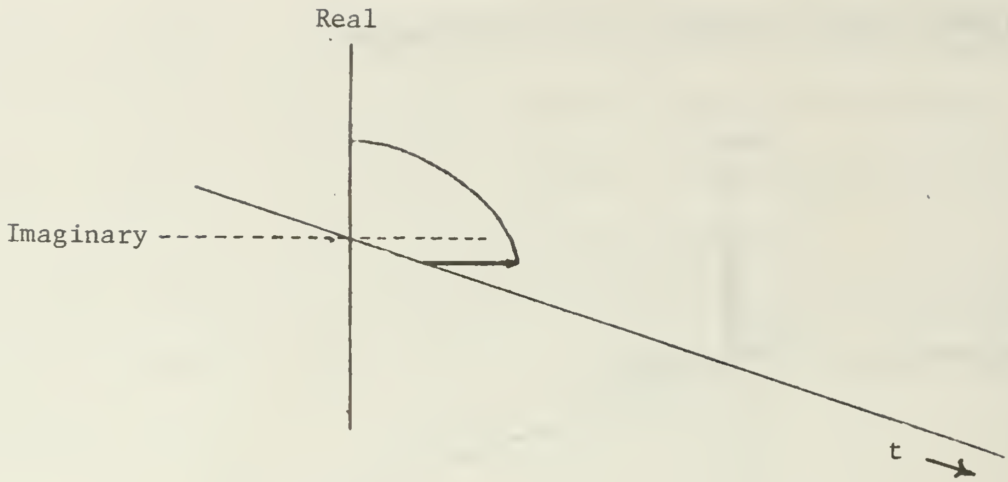


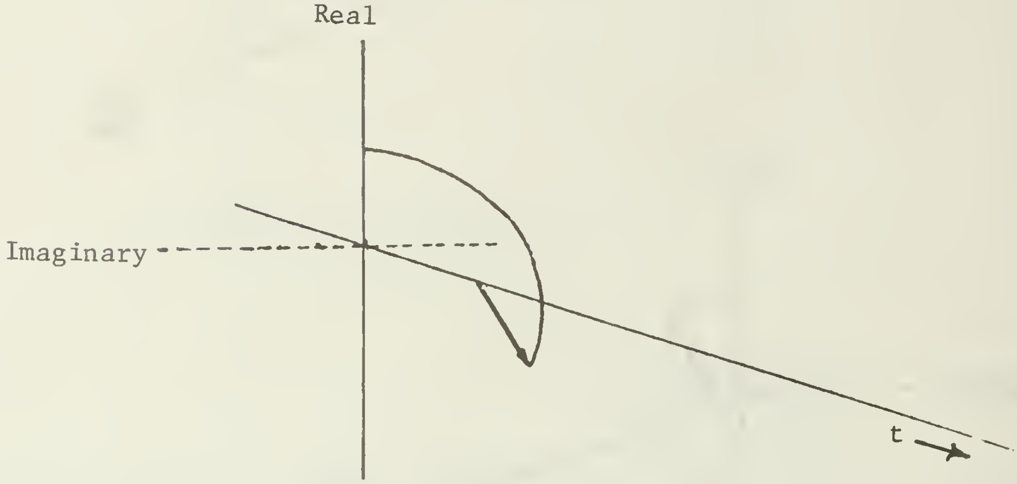
path around the time axis, as illustrated:

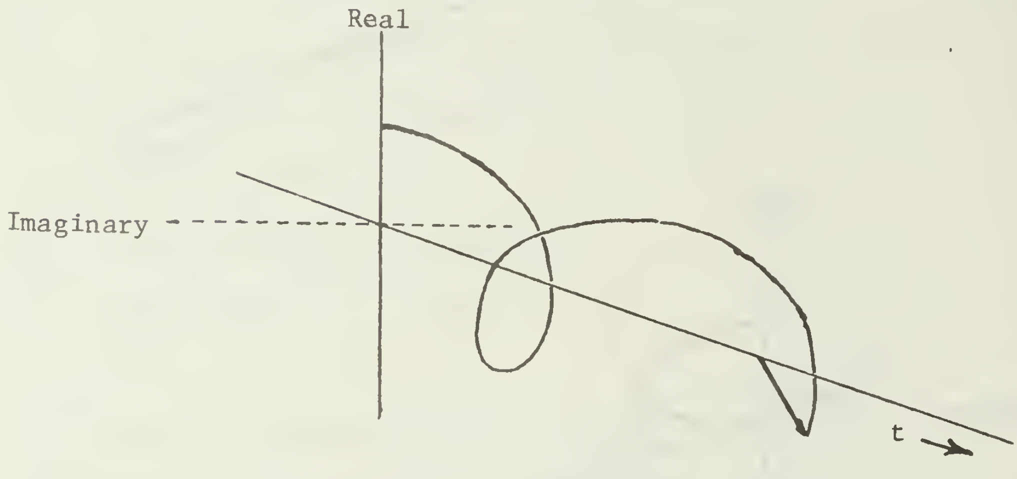








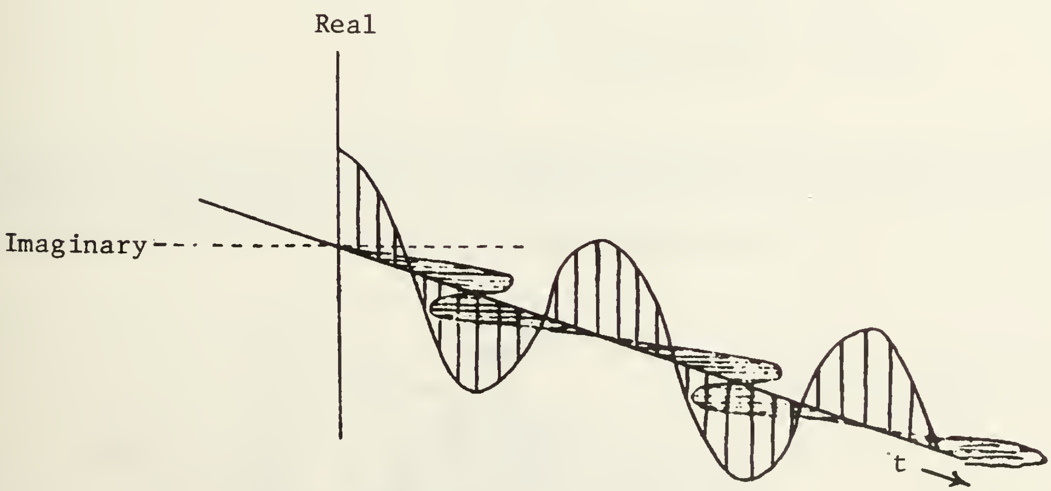




At this point it should be noted that there is another measure of angular rotation known as the "radian". There are 2π radians in 360°. This measure is used because it simplifies the mathematics involved in solving wave equations.

Returning to the vector generator, note that the number of times it makes a complete 360° (2π radians) sweep in one second of time is the frequency of the signal. The rotational velocity of the vector, ω, is the number of radians swept out by the vector per second. Thus, $\omega = 2\pi f$, since it sweeps out 2π radians each 360° rotation and makes f rotations per second. Thus ω is known as the "radian frequency" of the signal.

Now consider the projections of the vector on the real and imaginary planes as it rotates. The projection that appears on the real plane is a cosine wave and the projection on the imaginary plane is a sine wave. Also note that the vector rotates 360° during one cycle of the cosine wave on the real plane. This is where the use of 360° of phase angle arose.



The vector has components in the real and imaginary planes. The relationship connecting sinusoids and vectors is called the Euler (pronounced "oiler") Formula:

$$e^{\pm j\theta} = \cos\theta \pm j \sin\theta$$

where j is taken to mean "in the imaginary plane".

By vector addition, the signal is generated by vector-sum of the real component, $\cos \omega t$, and the imaginary component, $j \sin \omega t$.

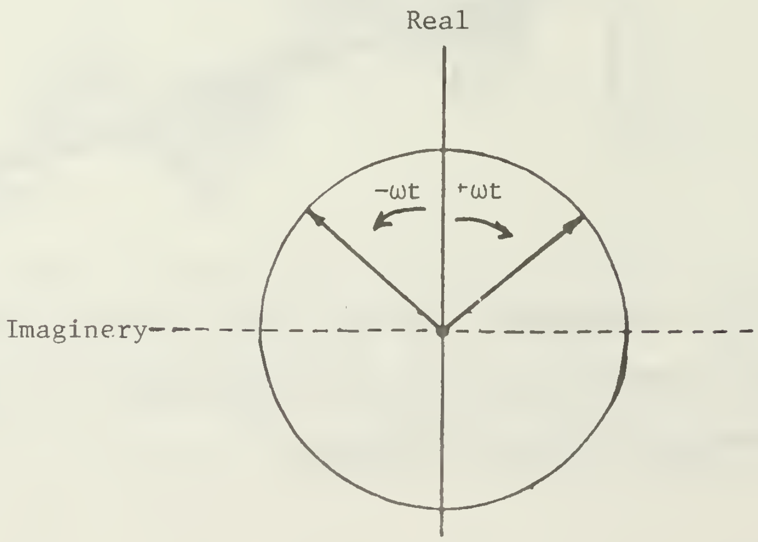
$$\text{Vector} = \cos \omega t + j \sin \omega t$$

Note that this is of the form of Euler's equation if $\theta = \omega t$. Thus, the mathematical description of the vector is $e^{\pm j\omega t}$, sometimes called a "phasor" because it varies only in phase.

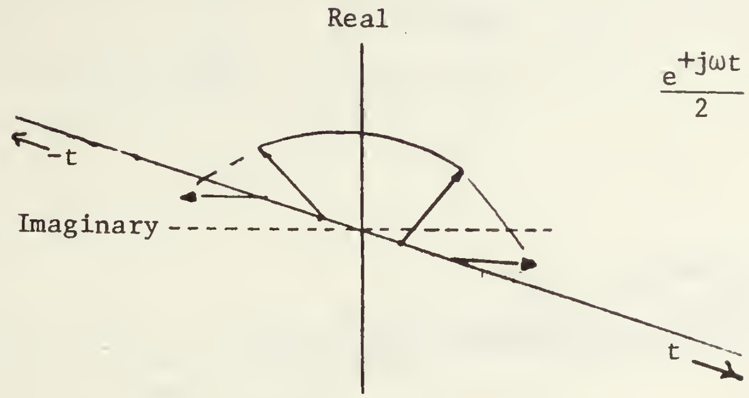
It was shown that $\cos \omega t$ is the projection of the phasor tip path on the real plane. Solving Euler's equation for $\cos \omega t$:

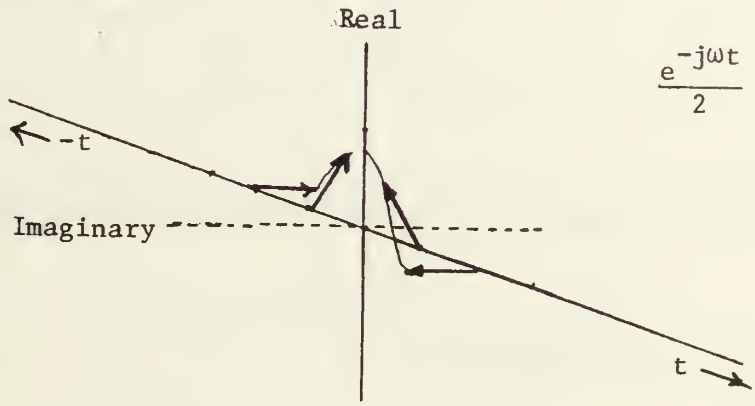
$$\text{Cos}\,\omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2}$$

Therefore a cosine wave is represented by two phasors rotating in the $+\omega t$ and $-\omega t$ directions, respectively.

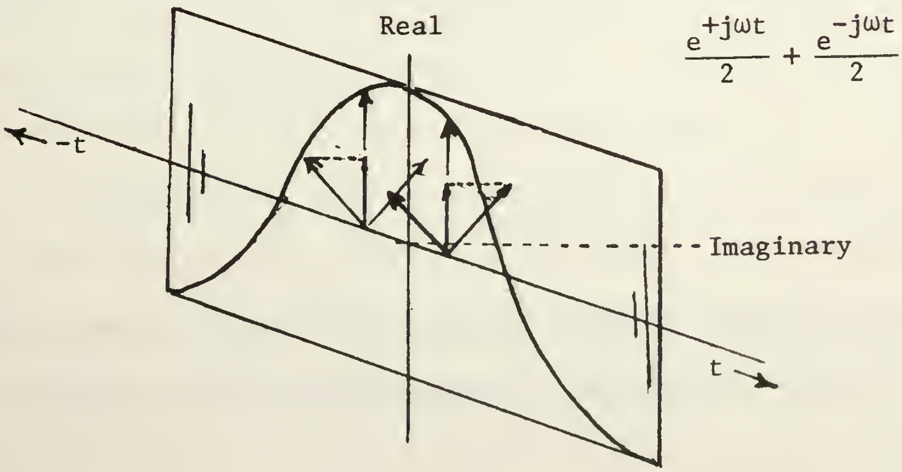


These phasors unwind in opposite directions along the t-axis.

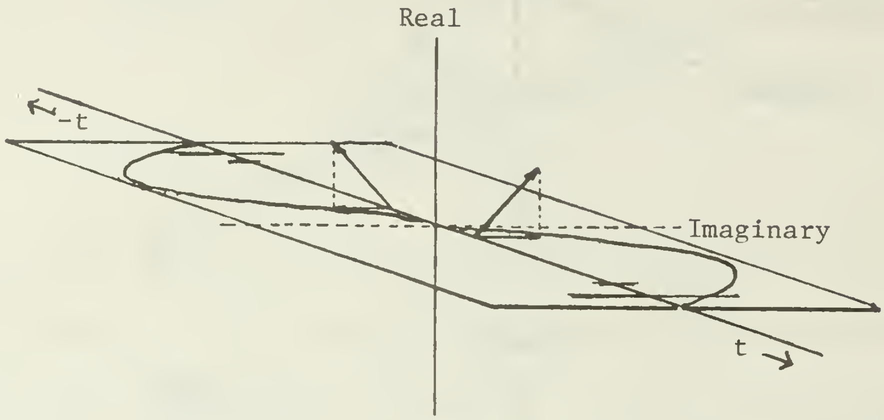




Looking at the projection of these phasors on the real plane, it can be seen that cos ωt is generated.



For the cosine wave, the two contributions of the phasors always add in the same direction on the real plane, therefore, the plus sign (+) in the formula.

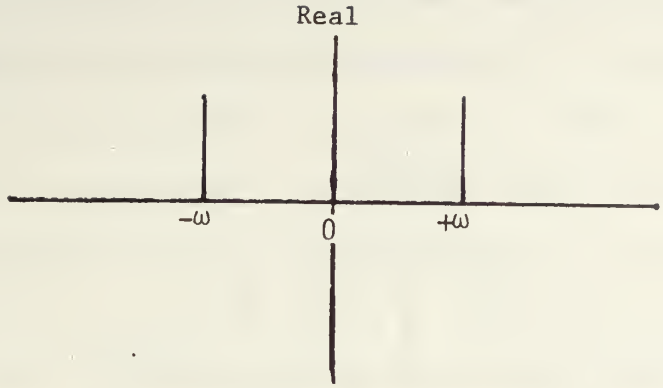


Solving Euler's equation for sin ωt, one obtains;

$$\text{Sin } \omega t = \frac{e^{j\omega t} - e^{-j\omega t}}{2j}$$

Because the two projections are always in opposite directions on the imaginary plane, there is a minus (-) sign in the formula.

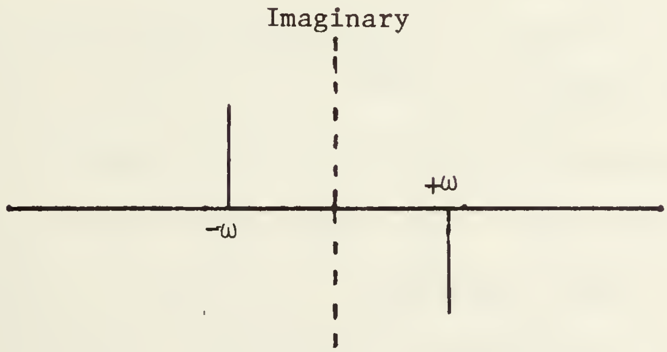
It was shown how a single frequency plots as a single line in the frequency domain. This appears as a line for each phasor. But note also that the mathematical solutions establish <u>two</u> phasors for each frequency, one rotating at +ω and one rotating at -ω. Thus, in actuality, there has been shown only one side of the spectrum, the positive frequency side. The full spectrum of cos ωt is as shown on the following page.



The lines are both on the plus side of the ω axis because of the plus sign in the formula:

$$\text{Cos } \omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2} .$$

The full spectrum of sin ωt is:



Note that the lines are on opposite sides of the ω axis due to the minus sign in the formula:

$$\text{Sin } \omega t = -j \frac{e^{+j\omega t}}{2} + j \frac{e^{-j\omega t}}{2}$$

Note also that the lines correspond to projections on the real plane for cos ωt, and the imaginary plane for sin ωt.

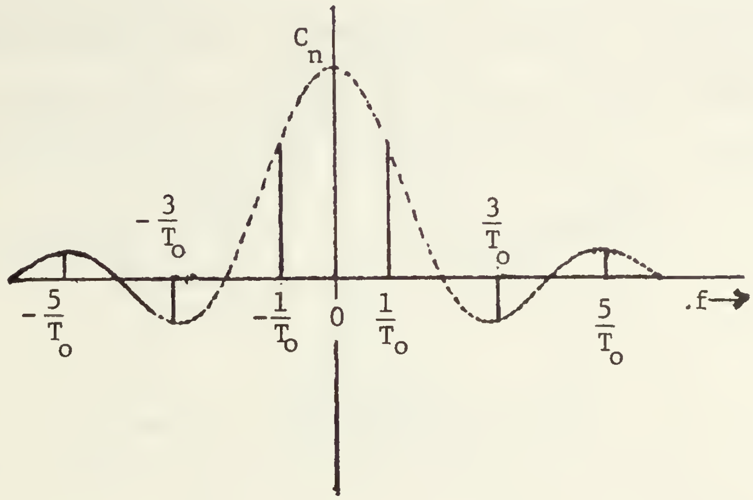
Fourier analysis shows that arbitrary waveforms can be made from combinations of sines and cosines of the proper amplitudes, phases, and frequencies. How can one determine which frequencies and what amplitudes and phases to use? In effect, one compares the waveform with various frequencies by overlaying the waveform on a given frequency to see how well they match. One measure of this is the point by point <u>product</u> of the two, <u>averaged</u> over a period. This measure, known as C_n (n^{th} frequency coefficient) is the following:

$$C_n = \frac{1}{T_o} \int_{-\frac{T_o}{2}}^{\frac{T_o}{2}} x(t)\, e^{-j\omega_n t}\, dt \qquad \omega_n = 2\pi n \frac{1}{T_o}$$

where ω_n is the n^{th} "harmonic" frequency component and $x(t)$ is the waveform of interest, $e^{-j\omega_n t}$ is the phasor being compared, and the integral serves to sum the product over the period T_o of the waveform. This determines how closely the waveform matches a given frequency. In order to find the components of a given waveform, the waveform must be compared with all n frequencies. The "spectrum" of the waveform is then the summation of C_n's multiplied by the n^{th} phasor component ($e^{j\omega_n t}$), known as the Fourier Series.

$$x(t) = \sum_{n=-\infty}^{\infty} C_n\, e^{j\omega_n t}$$

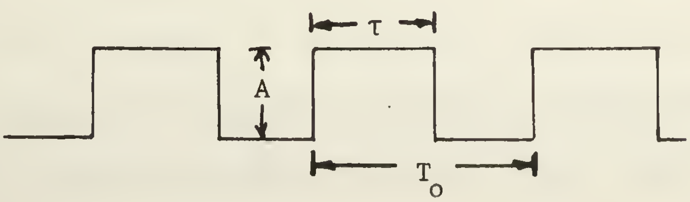
Using this technique to analyze a square wave train, the spectrum is obtained as shown below:



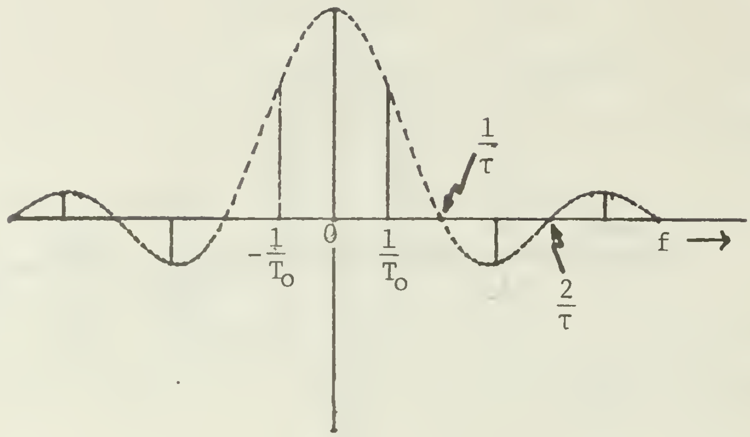
The fact that some of the C_n's are negative indicates that the frequency corresponding to that C_n is to be 180° out of phase with the positive sign C_n frequencies. In general, the C_n's may have any phase.

Note how all the C_n's occur at frequencies which are multiples or "harmonics" of the lowest component. This is because the signal is compared with sines and cosines which complete whole numbers of cycles during the period T_o. The ancient Greek, Pythagoras, discovered harmonics in the musical scale. Centuries later, the mathematician, Fourier, laid the theoretical foundation for signal analysis as described in this section.

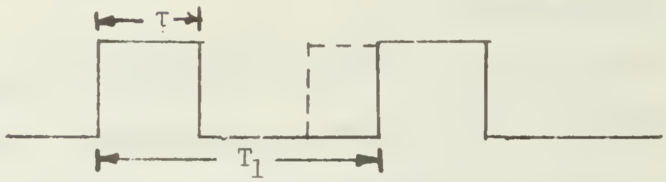
Now consider some square wave pulse trains and analyze their spectra:

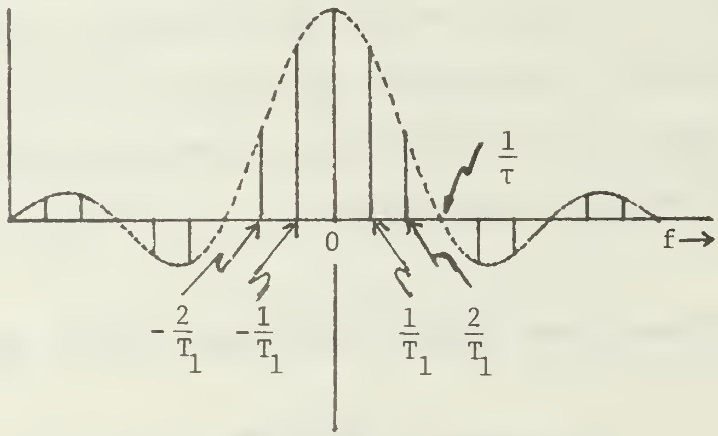


First solve for the C_n's and plot them. The curve is obtained as shown.



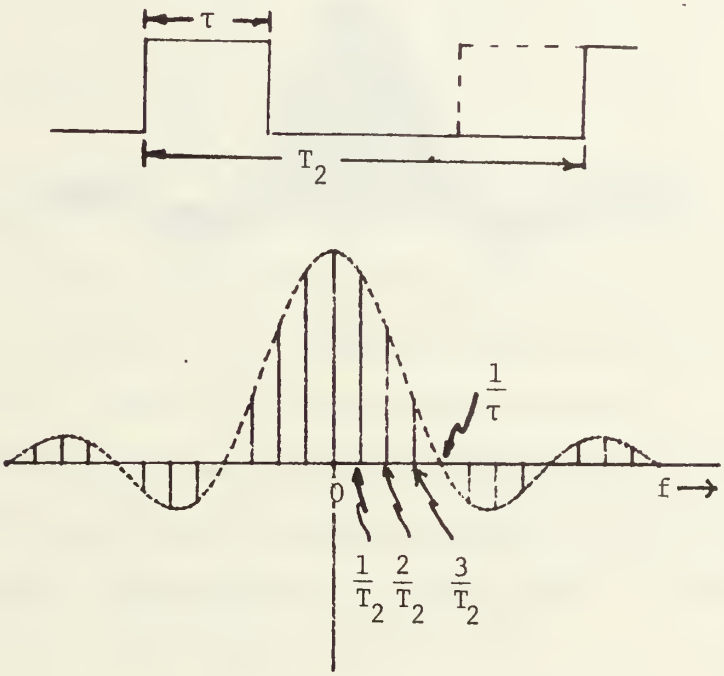
If the period of the pulse train is increased but the length of each individual pulse is kept constant, the following curves are obtained.





Note that the point where the envelope of the C_n's crosses zero remains the same, but the number of C_n's increases.

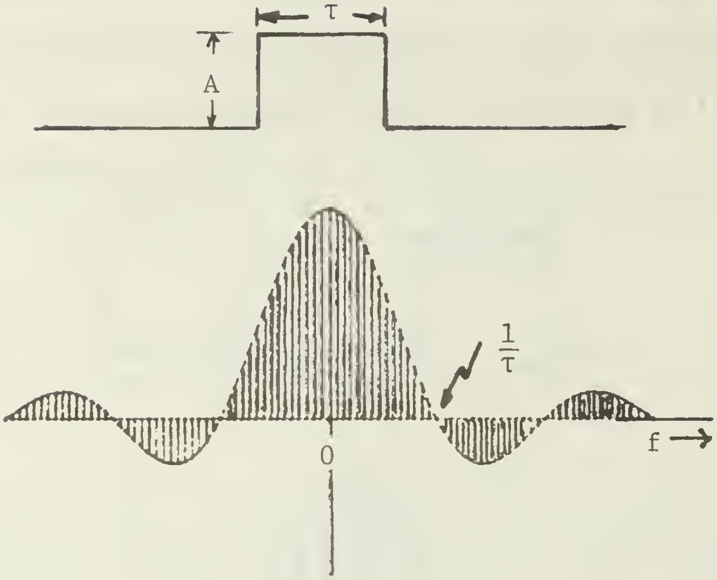
That is, there are more frequency components present in this signal than there were in the first. If one now increases the period even more, still keeping the pulse length constant, note how all the zero crossings of the envelope remain constant, but the number of C_n's increases again. If one were



to increase the period to the point where there were a single square pulse left, that is, $T \to \infty$, the result would be that the number of C_n's would be infinite, but the zero crossings of the envelope would remain the same, as illustrated on the following page.

From these examples we can see that the zero crossing interval of the envelope is related to the inverse of the pulse width, and the interval between C_n's is related to the inverse of the period of the signal. This, and other proper-

ties of the Fourier relations, will be discussed in greater detail in a later chapter.



The form of the envelope of the C_n's for the square wave appears so frequently that it has been given a name of its own. The form is (sin x)/x, which is called "sinc" (may be pronounced "sin-cee" to avoid confusion with "synch").

<u>Fourier Transforms:</u> It can be seen that as the period of a signal becomes infinite, the number of spectral components then becomes infinite. In this case, the spectrum of a signal can be characterized by the formula of the envelope of the C_n's. Techniques have been developed for finding the formula of the envelope in the limit as $T \to \infty$. One method of doing this is known as the Fourier Transform, defined as:

$$X(f) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt$$

Note that this is quite similar to the formula used for finding the C_n's. Instead of giving each individual C_n, however, this formula gives instead the envelope of the C_n's in the f-domain.

Now compute the transform of a single square pulse, using this formula. By integration the solution has the following form:

$$X(f) = [A\tau] \frac{\sin \pi\tau f}{\pi\tau f}$$

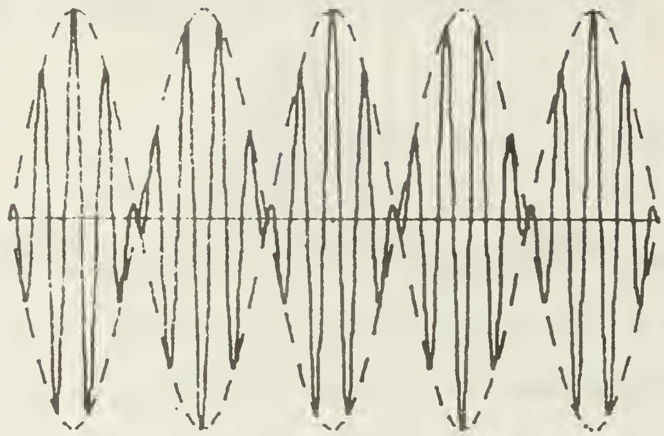
This is the same result obtained by computing the C_n's for a square wave of $T \to \infty$, solved for the formula of the envelope.

Since the spectrum, or f-domain characteristics of a signal can be found, given its time-domain formula, it is only reasonable that the reverse operation should be possible. This turns out to be the case, and the operation is accomplished by using the "inverse transform", defined as:

$$x(t) = \int_{-\infty}^{\infty} X(f)\, e^{j\omega t}\, df$$

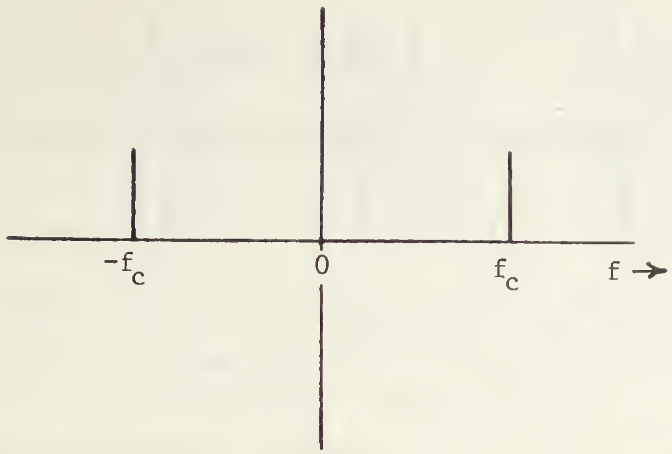
By substituting the spectrum in the inverse transform and integrating, the result is the time domain characteristics of the signal. The two transform formulas, forward and inverse, are called a "transform pair". One formula yields the spectrum, given the time-domain signal, and the other finds the time-domain characteristics, given the frequency spectrum. Later examples show how these are powerful tools used in signal processing.

For instance, returning now to the cosine wave that was introduced earlier, note what happens if one applies modulation. That is, impress a low frequency cosine wave on a "carrier" wave of a much higher frequency, f_c. A modulated cosine wave looks in the time domain as pictured in the following:

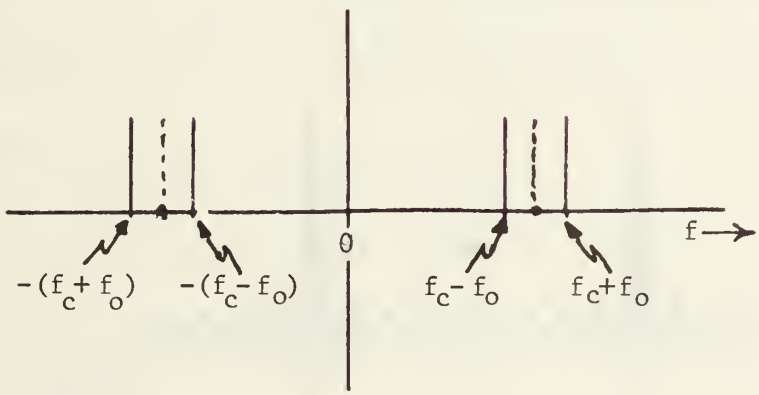


where the envelope varies as one cosine wave, while inside it are the individual peaks now at the carrier frequency. What does this signal look like in the f-domain? Given that the f-domain plot of the cosine wave looks as sketched:

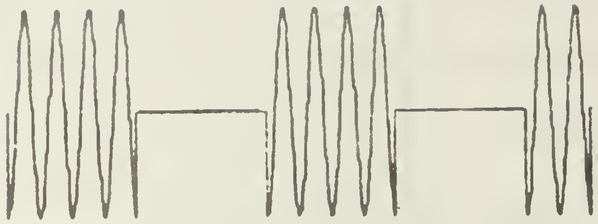
The carrier is just a cosine wave of a different frequency, so its plot looks as shown on the following page. Note that both f-domain plots were centered about f = 0.



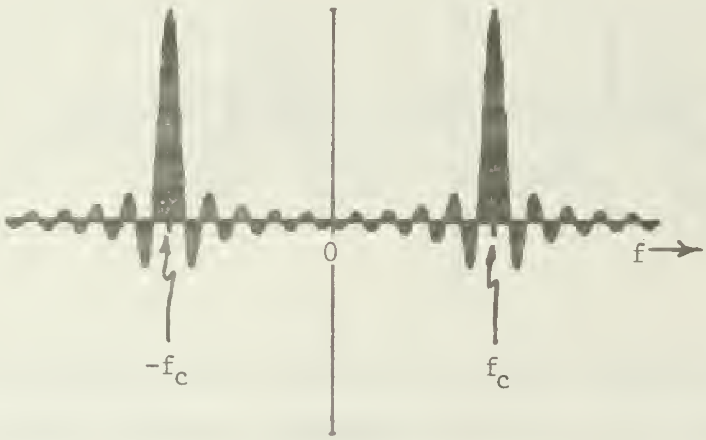
Now, modulating the carrier wave, i.e., multiplying it by the cosine wave, the following result is obtained by using the Euler formula:



Note that the resulting signal is of the form Cosine $(f_c \pm f_o)$. The f-domain plot is the cosine spectrum centered about the carrier frequency values instead of about f = 0. Thus, modulation in the time domain is equivalent to a translation of the spectrum in the f-domain from being centered at the f = 0 axis to centered about the carrier frequency. Another example: a square wave modulated carrier as illustrated on the following page.



It was shown that the spectrum of a square wave is a sinc function. Using the result that modulation of a signal translates the f-domain plot of the signal and centers it about the carry frequency, the f-domain plot of a square wave modulated carries is as shown. The envelope takes the shape of the square wave spectrum but is centered about the carrier frequency instead of about f = 0.



SELF TEST I

INTRODUCTION TO FOURIER TRANSFORMS

1. The fundamental concept of Fourier theory is that any arbitrary function of time x(t) can be decomposed into, or synthesized by, a summation of _______________.

2. The equation $X(f) = \int x(t)\, e^{-j2\pi ft}\, dt$ is called ________.

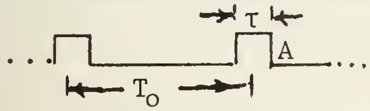
3. A phasor is _______________.

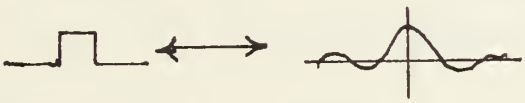
4. Continuous wave "CW" time functions are represented by ______________ in the frequency domain. Single pulses are represented by ____________ in the frequency domain.

5. Harmonics are _______________.

6. Sketch and dimension the graph of $\frac{\sin x}{x}$ (or Sinc x).

7. Sketch and dimension the spectrum of a train of pulses



8. The sketch  is called a

________________.

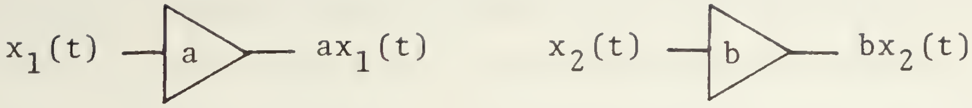
II. FOURIER TRANSFORM PROPERTIES

The properties of Fourier Transforms are stated most succinctly in the accompanying Fourier Transform Theorems. Knowledge of the theorems is important in the interpretation of spectra, as the theorems express the relationship between time-domain and frequency-domain operations.

In the introduction to Fourier Transforms, it was shown that the frequency spectrum X(f) of a signal x(t) was the Fourier Transform of the signal. That is, $X(f) = \mathfrak{F}[x(t)]$ and inversely, $x(t) = \mathfrak{F}^{-1}[X(f)]$. These relations are called "transform pairs" and are more conveniently denoted by $x(t) \longleftrightarrow X(f)$. The Fourier Transform Theorems describe the properties of various Fourier Transform pairs.

A. LINEARITY (OR SUPERPOSITION) THEOREM:

> If a signal $x_1(t)$ is multiplied by a constant, a (as would happen if an amplifier with a gain = a were placed in a circuit) and a signal $x_2(t)$ is multiplied by a constant b, then their respective line spectra will also be multiplied by the constant a or b.



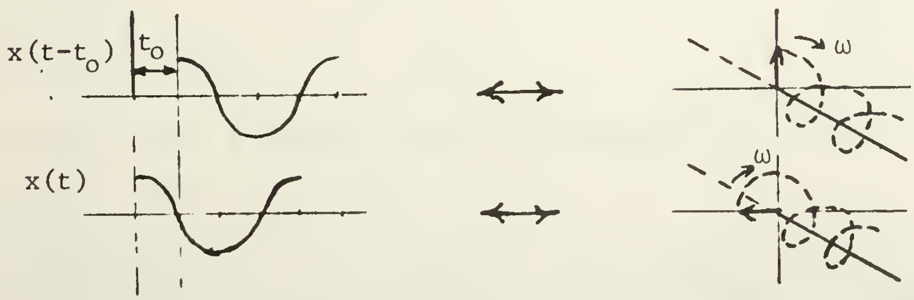
$$ax_1(t) + bx_2(t) \longleftrightarrow aX_1(f) + bX_2(f)$$

Linear combinations in the time-domain become linear combinations in the frequency-domain.

B. TIME DELAY THEOREM:

> Given the transform pair $x(t) \longleftrightarrow X(f)$ if the signal $x(t)$ is delayed by t_o seconds, as happens when the signal passes through an ideal delay line to become a new signal, $x(t-t_o)$, then the spectrum is modified by a frequency dependent phase shift to become $X(f)e^{-jt_o\omega}$

$$x(t-t_o) \longleftrightarrow X(f)e^{-jt_o\omega}$$



The translation of a signal in time changes the phase of the spectrum but does not alter the complex amplitude. To illustrate the Linearity and Time delay theorems, consider a signal which is a linear combination of delayed signals:

$$y(t) = 2x(t-t_o) - 4x(t-3t_o)$$

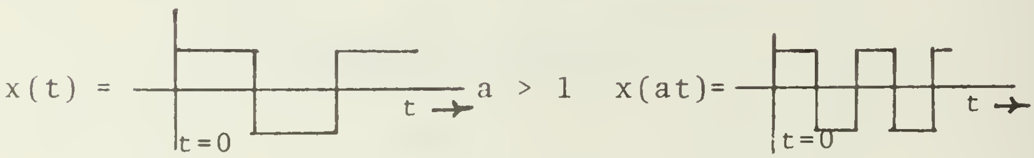
The spectrum is easily written as:

$$Y(f) = 2X(f)e^{-jt_o\omega} - 4X(f)e^{-j3t_o\omega}$$

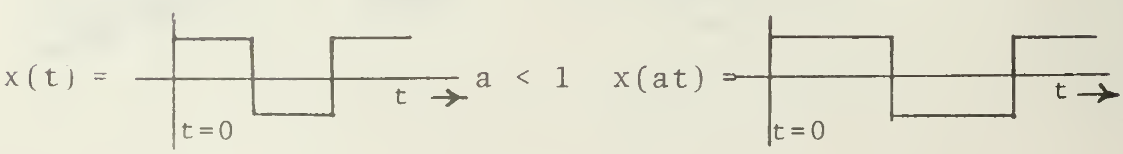
C. SCALE CHANGE THEOREM

It has been shown that a translation of the time origin may be accomplished using the Time Delay theorem. The time axis may also be expanded, compressed, or reversed by an operation known as "scale change". If a signal x(t) becomes a new signal x(at), then:

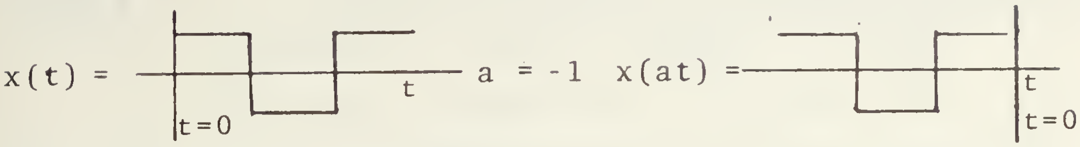
(1) x(at) is compressed if $\underline{a}$ is a number greater than one;



(2) x(at) is expanded if $\underline{a}$ is a number less than one;



(3) x(at) is reversed in time if a is negative.

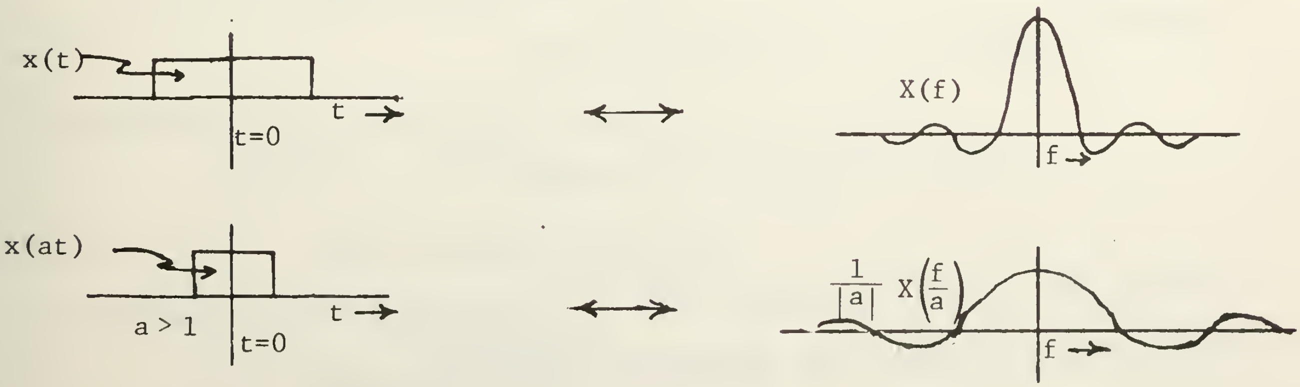


In the Fourier Transform pair,

$$\boxed{x(at) \leftrightarrow \frac{1}{|a|}\, X\left(\frac{f}{a}\right),}$$

$|a|$ means the absolute value of a, that is, the numerical value only is used and the sign is dropped.

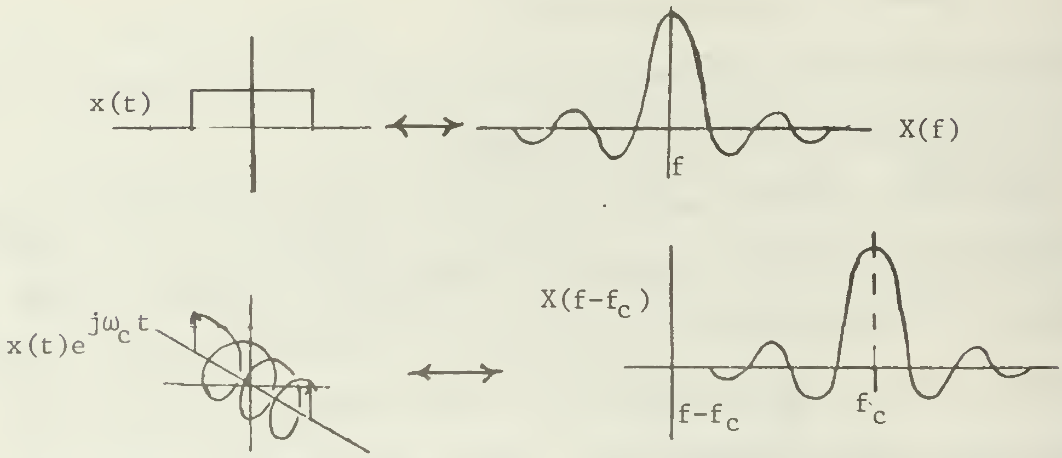
The Scale Change theorem expresses the property of reciprocal spreading. If the signal is compressed in time by the factor a, its spectrum is expanded in frequency by 1/a. The amplitude factor $\frac{1}{|a|}$ accounts for the change in area of the pulse due to time compression. The expansion or compression in the time axis of a signal occurs, for example, in the playback of recorded signals.



D. FREQUENCY TRANSLATION (MODULATION) THEOREM:

If a signal x(t) is multiplied by $e^{j\omega_c t}$, its spectrum X(f) will be translated in frequency by $\pm f_c$. Mathematically stated:

$$x(t)e^{j\omega_c t} \leftrightarrow X(f-f_c).$$

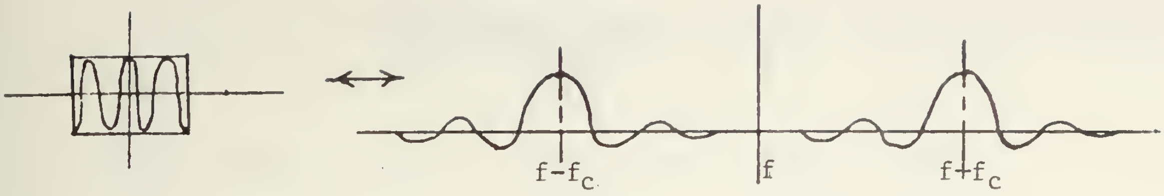


It is more common to multiply a time signal x(t) by the real part of $e^{j\omega_c t}$, i.e., by cos $\omega_c t$. From previous work with Euler's theorem, it is recalled that

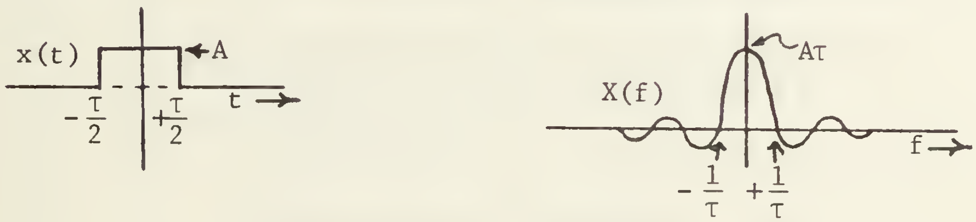
$$x(t) \cos \omega_c t = 1/2\, x(t) \left(e^{j\omega_c t} + e^{-j\omega_c t}\right).$$

Taking this result and applying it to the Frequency Translation theorem yields the Frequency Modulation theorem:

$$x(t) \cos \omega_c t \leftrightarrow 1/2\, X(f+f_c)$$



Or, if a signal x(t) is multiplied by cos $\omega_c t$, its spectrum X(f) is shifted up and down in frequency by an amount f_c. This is the same result noticed earlier when an RF pulse train was formed by multiplying a rectangular pulse train by cos $\omega_c t$. As noted then, and confirmed by the modulation theorem, time domain multiplication becomes translation in the frequency domain. As an example, a rectangular pulse of amplitude A, period t, centered at t = 0, and duration τ is $x(t) = A\Pi(t,\tau)$.

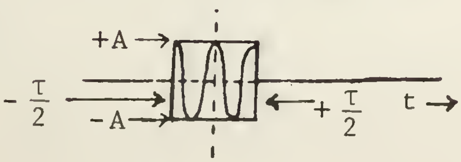


The spectrum is recognized as a SINC function, or

$$X(f) = A \text{ SINC } f\tau .$$

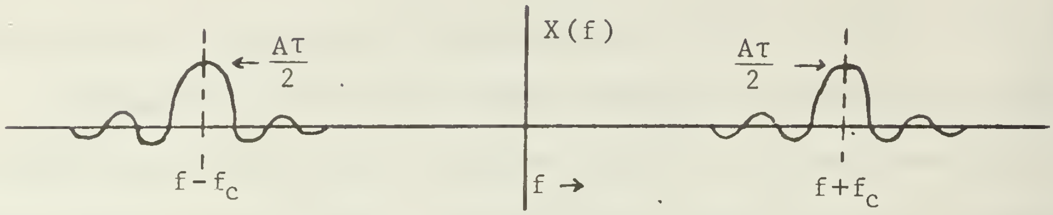
Then multiplying the single pulse by cos $\omega_c t$:

$$x(t) = A\Pi(t,\tau) \cos \omega_c t.$$



By the modulation theorem, the spectrum is

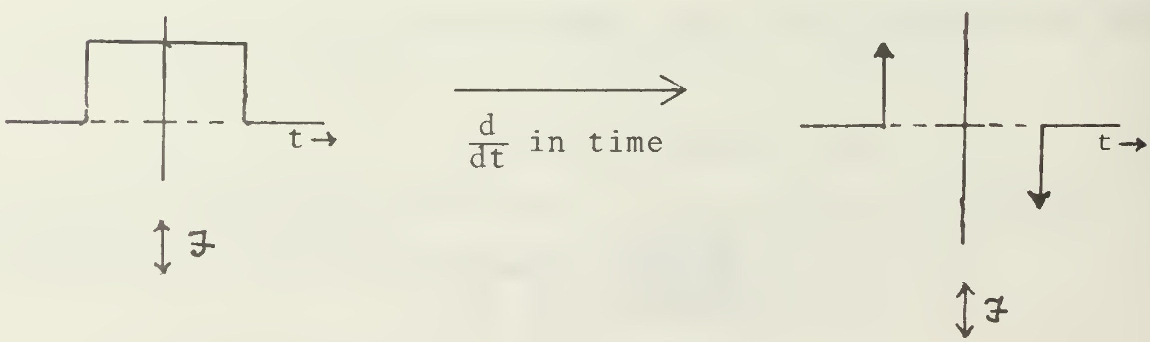
$$X(f) = 1/2\ A\tau\ \mathrm{SINC}\ (f-f_c)\tau + 1/2\ A\tau\ \mathrm{SINC}\ (f+f_c)\tau\ .$$

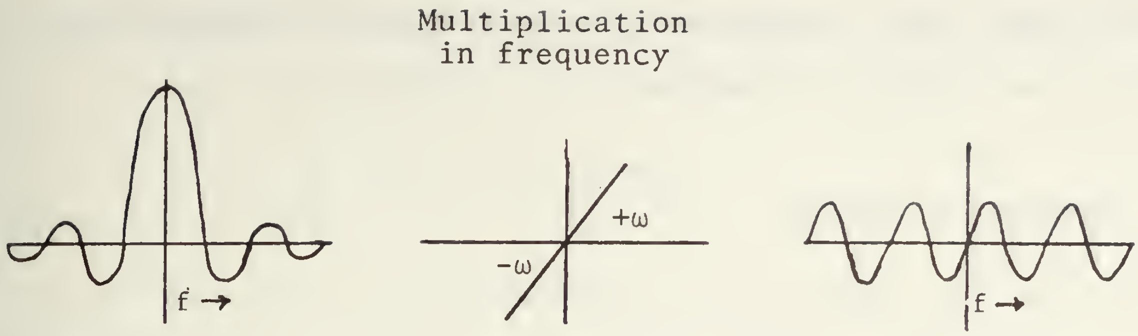


This is the result that is used extensively in single-side-band and double-sideband HF communications.

E. TIME DERIVATIVE THEOREM:

> If the derivative $\frac{d}{dt}$ is taken of a signal $x(t)$, the result is multiplication in the spectrum by $j\omega$.
>
> $$\frac{d}{dt}\,x(t) \leftrightarrow j\omega X(f)$$

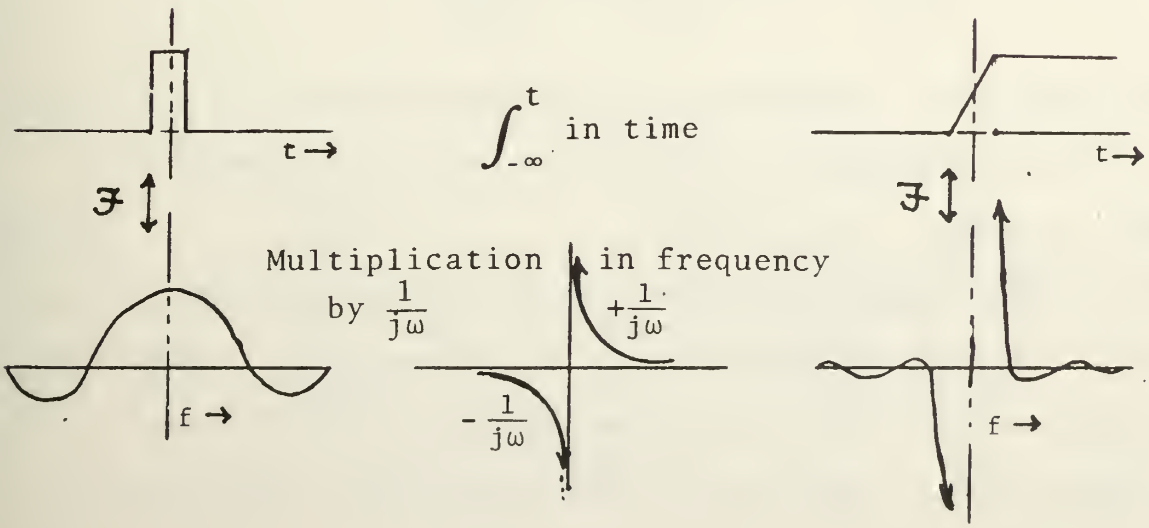




Therefore, differentiation enhances the high frequency components in the spectrum.

F. TIME INTEGRATION THEOREM:

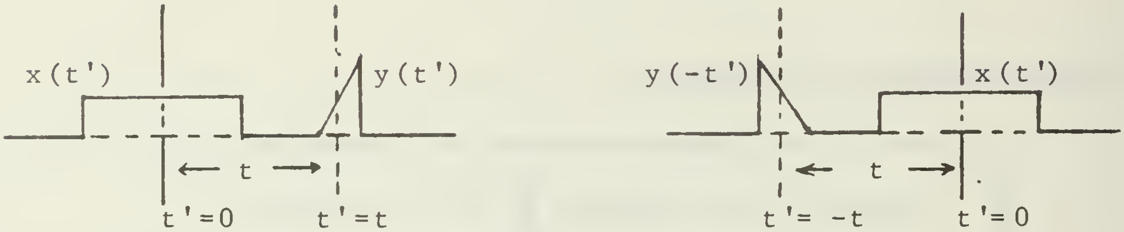
> Taking the integral $\int_{-\infty}^{t}$ of a signal $x(t')$ is equivalent to multiplying by $\frac{1}{j\omega}$ (dividing by $j\omega$) in the frequency domain:
>
> $$\int_{-\infty}^{t} x(t')\,dt' \leftrightarrow \frac{1}{j\omega}\,X(f).$$



Integration, then, suppresses the high frequency components but the low frequency components are unaffected.

G. CONVOLUTION THEOREM

Taking two signals in the time domain t', for example, x(t') and y(t'), AND "flipping" the y(t') signal to the other side of the axis, it becomes y(-t'), as illustrated below.



In order to make a comparison between the two signals, they are multiplied at every point in time t' from $-\infty$ to $+\infty$. Mathematically, this can be expressed as

$$z = \int_{-\infty}^{+\infty} x(t')\,y(-t')\,dt' .$$

In the case above, anywhere in the t' domain where x(t') exists, $y(-t') = 0$, and where y(-t') exists, $x(t') = 0$. Therefore, their product is zero.

Since the signals are separated by time t, if a time delay of t is introduced making y(-t') become y(t-t'), the signals will overlap. If t, the time delay, is varied from $-\infty$ to $+\infty$ the y(t-t') signal will "slip" across the signal x(t').

The result is expressed as a function of time t;

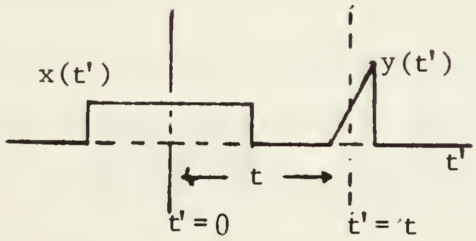
$$z(t) = \int_{-\infty}^{+\infty} x(t')\, y(-t' + t)\, dt'$$

The integral above is known as the "convolution integral" and expresses convolution mathematically, i.e., the area of the product of the two signals. A more convenient notation for convolution is:

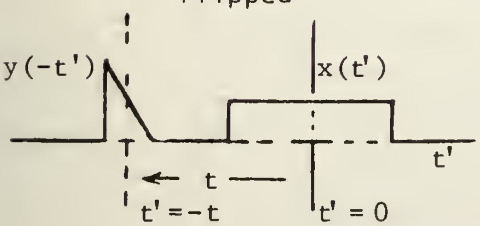
$$z(t) = x(t) * y(t).$$

A graphical interpretation of convolution, as shown may help to see what happens to the signals.

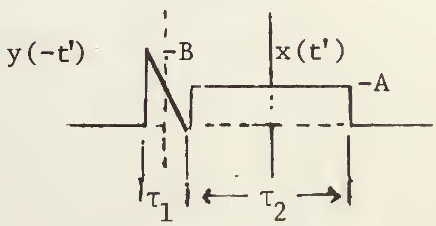
Two signals



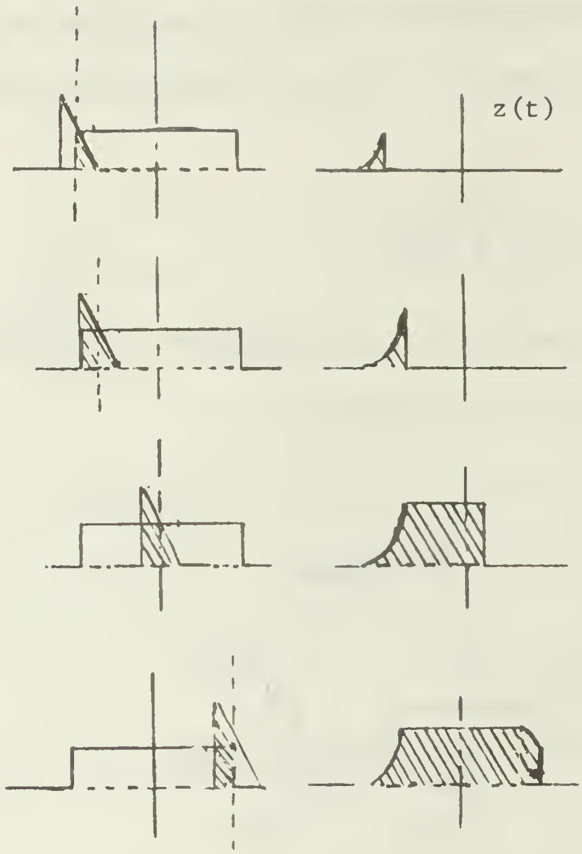
"flipped"



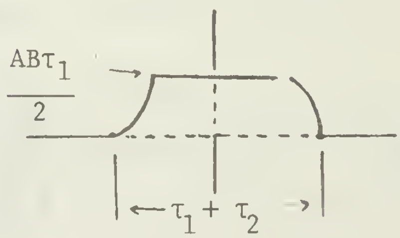
"slipped"



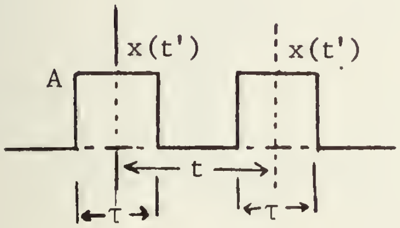
The shaded area on the time plot of both signals represents the area of their "common product" for a particular value of time delay t. The shaded area on the z(t) plot represents the summation of the area of their "common product" as the "flipped" signal y(-t') is "slipped" across x(t').

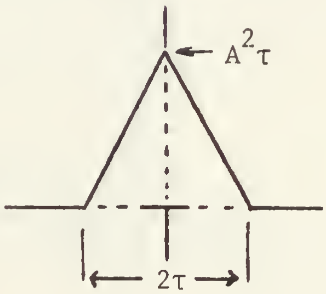


$$z(t) = x(t) * y(t)$$



The convolution (the "flip" and "slip") of two rectangular pulses produces a trapezoidal pulse with a base of length $\tau_1 + \tau_2$. Convolution of two identical rectangular pulses produces a triangular pulse with base equal to 2τ and height equal to $A^2\tau$.





$$z(t) = x(t) * y(t)$$

In this instance,

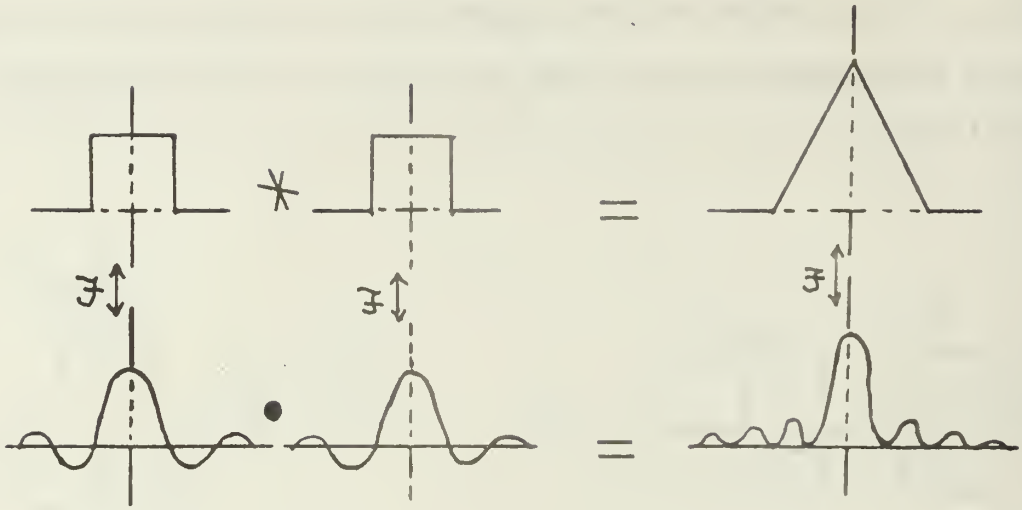
$$z(t) = x(t) * x(t).$$

Now that convolution in the time domain has been explained, how does it relate to the frequency domain? There are two convolution theorems which relate the two domains.

$x(t) * y(t) \leftrightarrow X(f) \cdot Y(f)$
Convolution in the time domain transposes to multiplication in the frequency domain.

$x(t) \cdot y(t) \leftrightarrow X(f) * Y(f)$
Multiplication in the time domain transposes to convolution in the frequency domain.

As an example of this, the now familiar square pulse is shown.



H. CORRELATION FUNCTION

In statistics, when it is desired to know how closely one distribution resembles another, the two are overlayed and the areas in common are multiplied. This "common product" represents the degree of similarity, or how well one distribution correlates to the other. In signal processing, it is often necessary to compare one signal with the same signal which occurs later in time or with another signal. Signals vary in time in different ways and the time origin may not be known. Therefore, in order to overlay one signal on another completely, the signal is displaced in time by τ units, then τ varied to "slip" the displaced signal across the one with which it is desired to correlate. With each τ used, the "common product" will be a different value. In order to obtain the value of correlation at a given delay τ, the "common product" is

summed over all values of time in an interval T and divided by time T (or multiplied by $\frac{1}{T}$). This defines the "correlation function" $R(\tau)$ and is described mathematically as:

$$\boxed{R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} x(t)\ x(t+\tau)\ dt.}$$

This is known as a "time average" and is denoted by

$$R(\tau) = \langle x(t)\ x(t+\tau)\rangle$$

Note that in this context, τ is a time "slip" not a pulse length. If the signal is periodic in T_o it need only be averaged over one period T_o instead of all time. The correlation function then becomes:

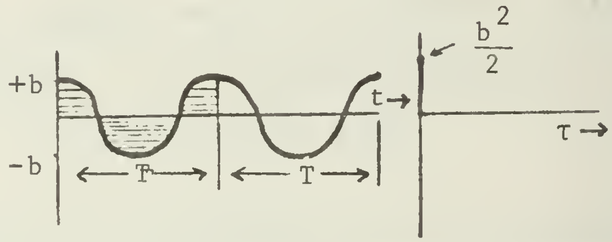
$$\boxed{R(\tau) = \frac{1}{T_o} \int_{-T_o/2}^{+T_o/2} x(t)\ x(t+\tau)\ dt\ .}$$

$R(\tau)$ is more precisely referred to as the autocorrelation function, that is, the correlation of one signal with itself. It may be desirable, as mentioned earlier, to correlate one signal $x(t)$ with another signal $y(t)$, displaced τ units in time to become $y(t+\tau)$. This results in the cross correlation function, $R_{xy}(\tau) = \langle x(t)\ y(t+\tau)\rangle$. The cross-correlation again measures the similarity. Where "correlation function"

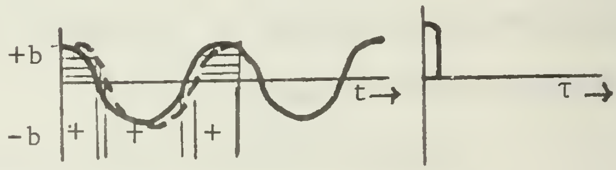
is used, it is assumed to be the "autocorrelation function".

For a graphical representation, a sinusoidal wave is used which, as illustrated, represents the result of the summation of the "common product" as the delayed wave is "slipped" across the other wave. This part of correlation is like the previously described convolution except that the signal is not flipped and the product is divided by T.

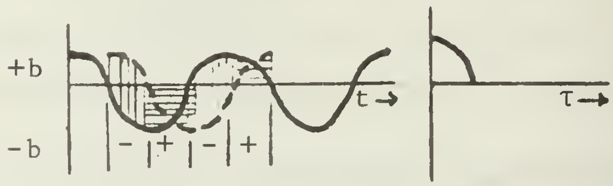
When both signals overlay completely, the value is large and positive.



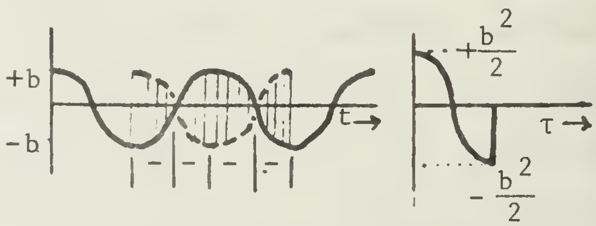
A small shift will give some negative area but the result is still positive.



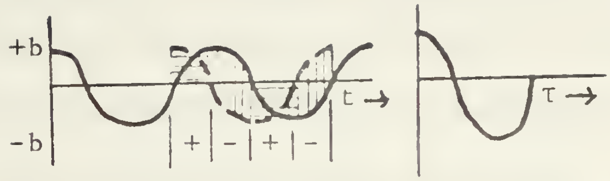
A 90° shift makes the positive and negative areas equal and the resultant value zero.



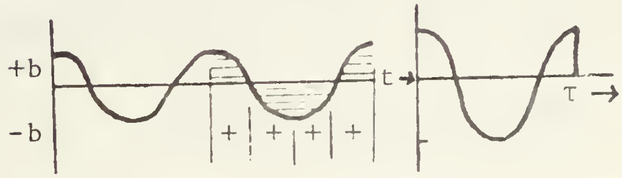
A shift of 180° results in the entire area being negative.



At 270° the positive and negative areas are again equal and the resultant value is zero.



At 360°, a complete one cycle shift, the signal is back in phase and the result is a maximum positive value.



The correlation function of a periodic signal is also periodic. Since the time average is taken, the maximum value of the autocorrelation is, in fact, equal to the average power of the signal.

Up to now there have appeared many similarities between convolution and correlation. However, if an asymetrical signal is used, the final result obtained by convolution may be quite different from the final result obtained by correlation. In addition, correlation theoretically includes a scale factor $\frac{1}{T}$ but in practice, T is finite and the distinction between correlation and convolution is essentially related to the "flip".

I. POWER SPECTRUM

An important theorem which applies to correlation is the fact that the Fourier Transform of the autocorrelation is equal to the power spectrum G(f).

$$\boxed{R(\tau) \leftrightarrow G(f)}$$

For periodic signals, the power spectrum is the <u>square</u> of the C_n coefficient magnitudes of the Fourier Series. It will be shown later that this same theorem applies to random signals for which the power spectrum is a continuous density function rather than discrete lines. There also exists a cross-spectrum for cross-correlation.

J. CONSERVATION OF ENERGY

It is reasonable to expect that a signal pulse viewed in the frequency domain should represent the same energy as in the time domain. Energy is given by integrating the power over all time. For signal x(t), the power is given by $|x(t)|^2$ (assuming for simplicity, one Ohm of resistance). Therefore, the energy is given by:

$$\boxed{E = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt}$$

The theorem states that energy is also given by:

$$\boxed{E = \int_{-\infty}^{+\infty} |X(f)|^2 \, df}$$

A similar theorem applies to average power $\overline{P}$ of continuous signals (which have infinite energy).

$$\boxed{\overline{P} = \langle |x(t)|^2 \rangle \int_{-\infty}^{+\infty} G(f) \, df = R(\tau=0)}$$

Note the connection to autocorrelation with zero time shift.

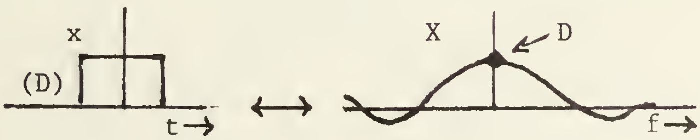
K. SIGNAL IMPULSE

In the study of mechanics, the quantity known as impulse is given by integrating force over all time. The electrical equivalent for a signal pulse is equal to the pulse <u>area</u> D in the time domain.

$$\boxed{D = \int_{-\infty}^{+\infty} x(t)\, dt}$$

Note that this is just the Fourier Transform evaluated at zero frequency (which causes the exponential to be unity). Therefore:

$$\boxed{D = X(f = 0)}$$



SELF TEST II

FOURIER TRANSFORM PROPERTIES

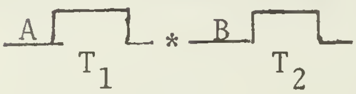
Explain in words:

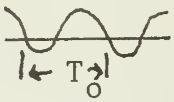
1. $a\ x_1(t) + bx_2(t) \leftrightarrow ?$

2. $x(at) \leftrightarrow ?$

3. $x(t)\ e^{j2\pi f_c t} \leftrightarrow ?$

4. $x(t)*y(t) = \int [\ ?\ ][\ ?\ ]dt \leftrightarrow ?$

5.  = ?

6. If $x(t) =$  , sketch $R(\tau)$

7. Compare $R_{xy}(\tau)$ with $x*y$

8. $R(\tau) \leftrightarrow ?$

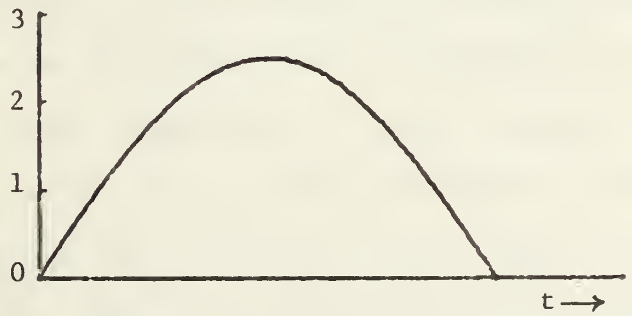
9. $\int_{-\infty}^{+\infty} |X(f)|^2\, df = \ ? \ = \int [?]\, dt$

10. Pulse area in the time domain = ?

III. QUANTIZATION AND SAMPLING

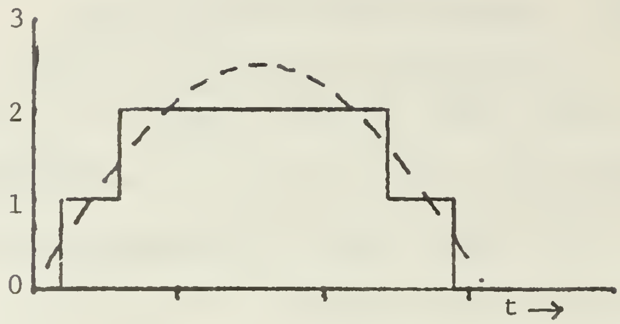
In the real world, most of the signals of interest to an operator of a detection system are of a continuous, or analog, nature. On the other hand, most of the more sophisticated processors in use today are digital in operation, and thus can not use analog inputs directly. Some means of digitizing the analog signals must be used.

To quantize a signal, the value of the signal at any given time must be described by a digital number. This requires a measurement and conversion of the signal value to digital units. The question of how accurately the value of the signal must be measured in converting it to digital form must be considered. Take the following signal, for example:

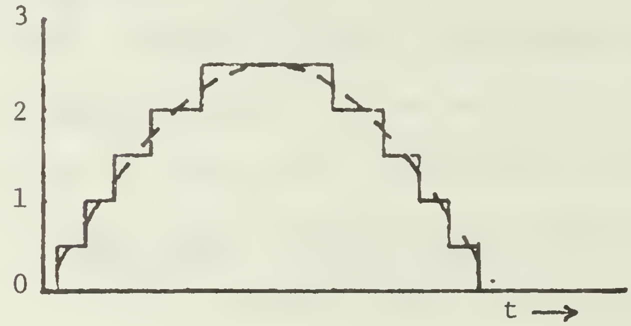


This signal has a maximum value of 2.5 and a minimum value of 0. Since the continuous values of the signal must be converted to a set of discrete values, the decision must be made of how many discrete values are needed in the range from 0 to 3 to describe the signal adequately. If it is decided to use four levels, that is, digitize to the nearest integer value, the

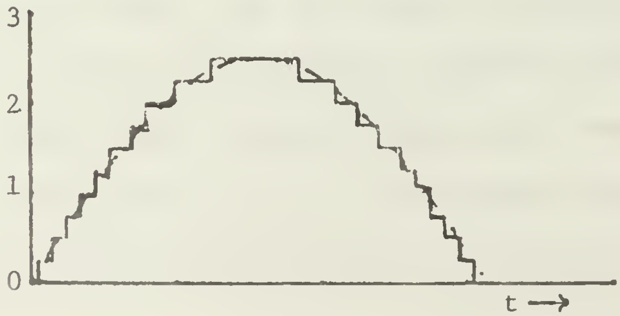
signal will have the following appearance:



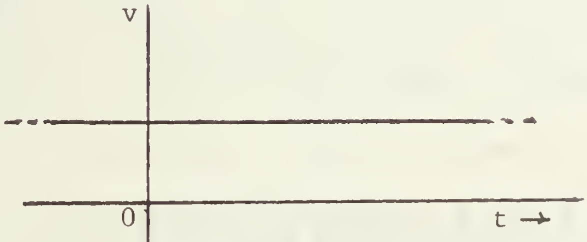
If the number of values is doubled and digitized to the nearest half integer value, the digitized signal looks like this:



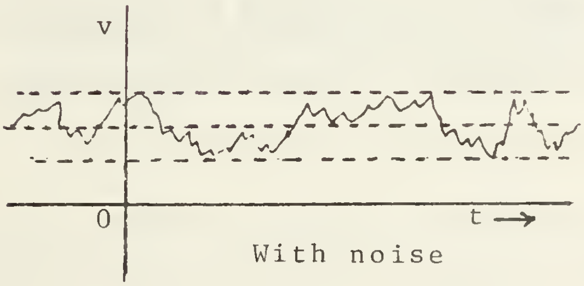
Note that as the number of values used in digitizing the signal increases, the digitized curve more and more closely approximates the original signal. The values used in digitizing the signal are known as quantization levels. In the case of a noise-free signal, the signal can be digitized to any degree of accuracy by increasing the number of quantization levels.



The case of a signal with noise is different, however, In this case, there is a random fluctuation of the signal value because of noise, even when the value represented by the signal itself is unchanging. The range of these noise fluctuations determines how accurately the value represented by the signal can be measured, even with a perfect measuring device. To illustrate this, consider first a signal without noise and then one with noise.



Without noise



With noise

It is observed that in the noisy signal, the value of the signal fluctuates about some mean, or average, value within some limits. If the noise is truly random in nature, the mean value about which it fluctuates is then the actual value of the signal being represented. At any given instant, any value of the signal between the upper and lower limits of the noise fluctuation could stand for the actual value being represented. The accuracy with which this signal can be measured is then

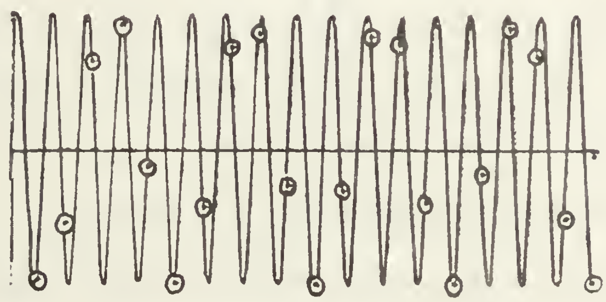
limited to the range of fluctuation due to noise. For example, suppose the signal of interest is a constant 3.0 volts but because of noise, the signal fluctuates between 3.5 and 2.5 volts. The most accuracy with which this signal can be measured is to the nearest integer volt, an appropriate degree of quantization.

Once it is determined how many quantization levels are necessary, it must then be determined how often to <u>sample</u> the signal in order to reconstruct it accurately. Obviously, if the signal never changes value but remains constant, only one sample is needed to reconstruct the signal for all times. If, however, the signal is not constant, there is a real problem.

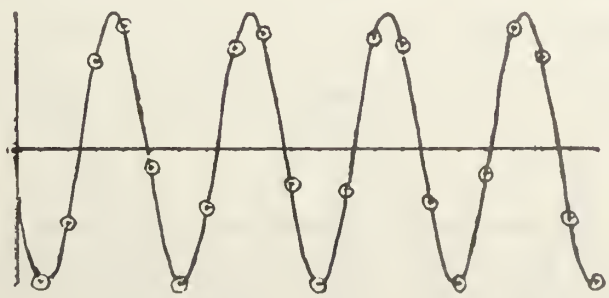
How often is it necessary to sample a given signal? It is known that one sample is not enough if the signal is not constant and also that if the signal is sampled continuously (i.e., in an analog fashion) it can be reproduced exactly. To sample a signal at extremely short intervals is very costly and difficult to accomplish. What is needed is a compromise, a sampling rate high enough that the signal can be reproduced from the samples but not any higher than is necessary in order to keep cost and equipment complexity down.

It is known intuitively that there must be some relationship between the rate of change of the signal and the sampling rate necessary to be able to reproduce it. If the signal changes very slowly, sampling need be done infrequently. If it changes rapidly, however, it must be sampled frequently enough that it does not have time to make several changes in

the interval between samples or the changes will be lost and the signal cannot be reproduced. If the sample intervals are too long, the signal is said to be "undersampled". This condition, in addition to causing the loss of some of the information in the signal, can also cause other complications. In some of the old cowboy movies the stagecoach wheels appear to turn backwards at times. This is a case of undersampling which illustrates one of the other effects, called "aliasing". To explain this, a sine wave is diagrammed.

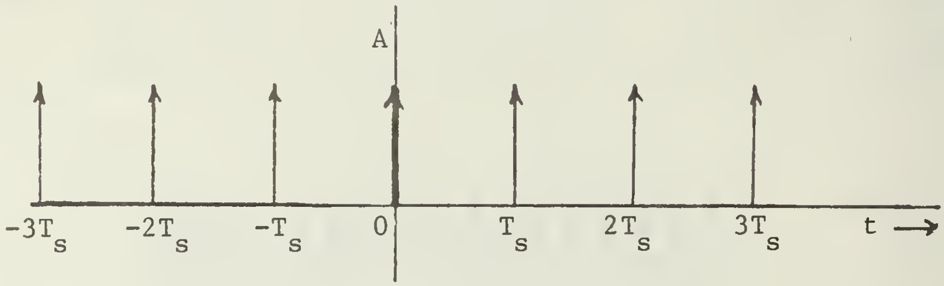


If this is sampled at a frequency less than that necessary for reproduction of the signal, a series of samples which give the impression of a sine wave of lower frequency is received.



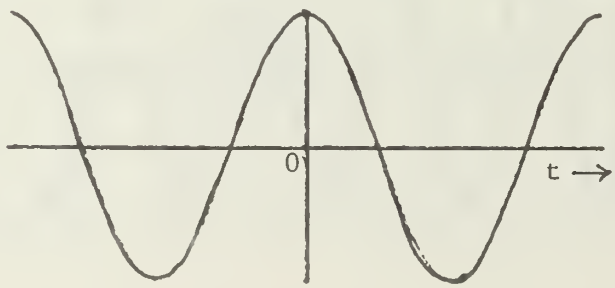
This is a problem which must be addressed when analyzing unknown signals for frequency components, since aliasing can give spurious results if the analyzer sampling rate is not high enough. This will be discussed further.

In the time domain, the sampling process can be visualized ideally as the multiplication of the signal by a function known as the "ideal sampling wave", which is identically equal to zero except at the sample times (T_s, $2T_s$, etc.).

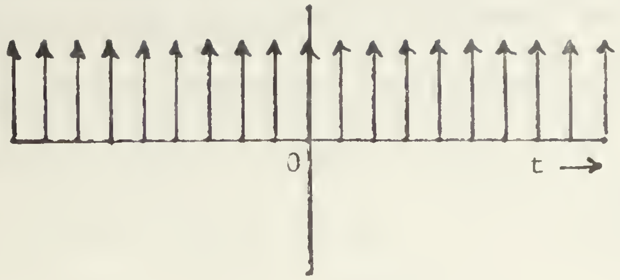


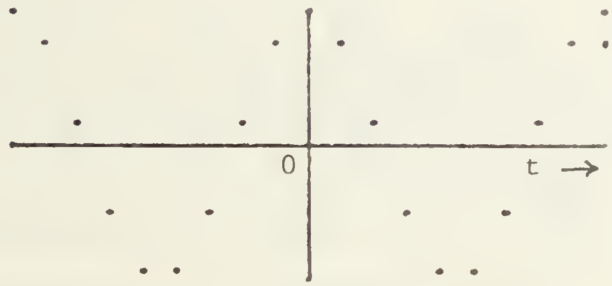
Ideal Sampling Wave

The duration of these spikes, known as "delta functions", is infinitesimal approaching zero, while their amplitude approaches infinity. They are defined such that their area ($0 \times \infty$) is equal to one. When the signal of interest is multiplied by the ideal sampling wave, the product of their areas is at each point T. Since the area of each delta function is one, and its width is zero, the product of areas is equal to the value of the signal at that time. Thus the products of the two functions appear as

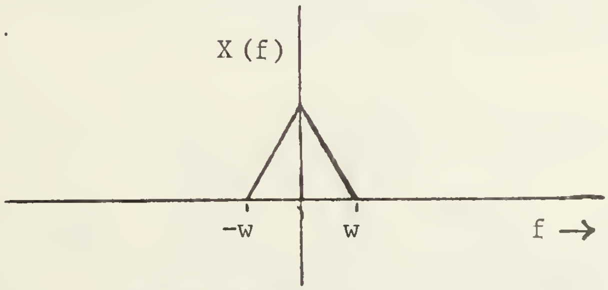


and

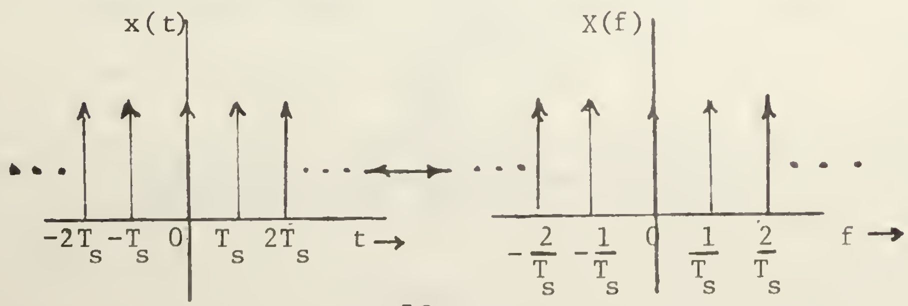




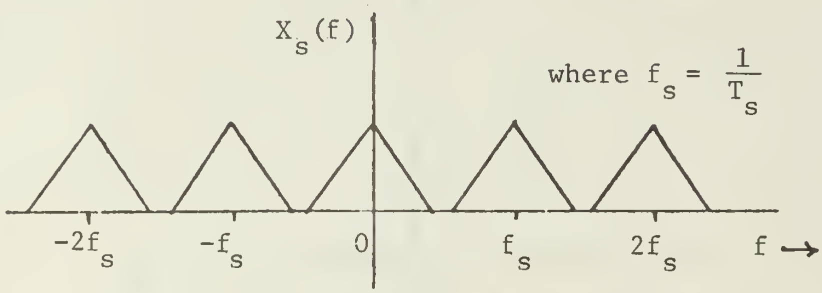
The frequency domain diagram indicates with more clarity what happens when a signal is sampled. Assume that the signal to be sampled is strictly "bandlimited", that is, it has no frequency components outside some limit.



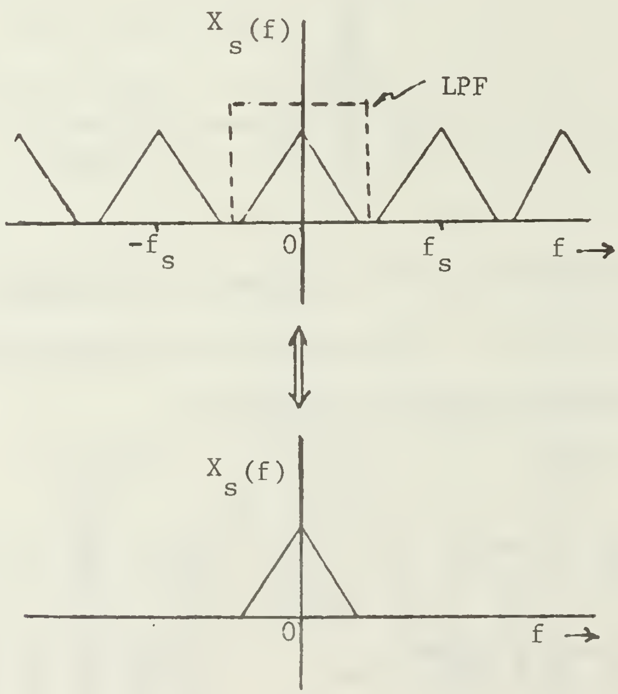
2W is the bandwidth of this signal. Although the mathematics is rather involved, it can now be shown that the ideal sampling wave spectrum appears as below.



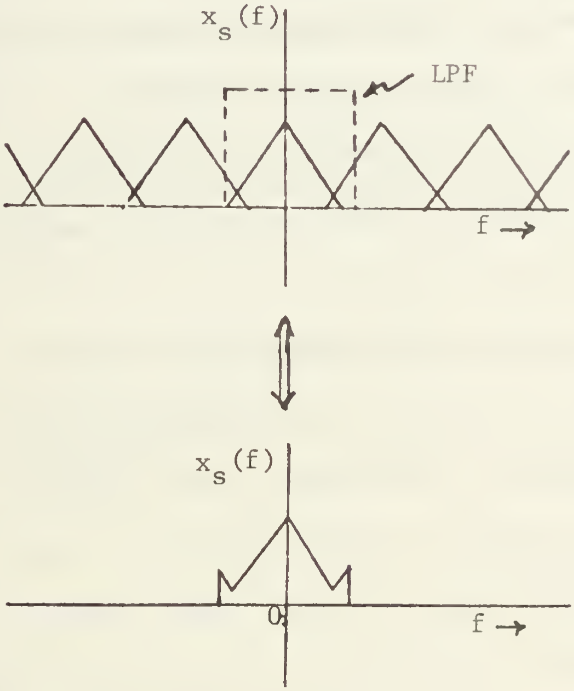
From the modulation theorem it is recognized that modulation of a function causes its spectrum to be shifted and centered about the modulation frequency and otherwise does not change it. Thus, the spectrum of the sampled signal appears as follows:



From this it is seen that for strictly bandlimited signals, if $f_s \geq 2W$, there is no overlapping of the spectrum components and the spectrum of the original signal is exactly reproduced. If the sampled signal is then passes through a low-pass filter, which has the property of allowing only those frequencies below its cutoff to pass, the original signal can be exactly recovered. (Filters are discussed in Section IV).



It follows from the previous diagram that the sampling frequency must be $\geq 2W$, i.e., greater than, or equal to, two times the highest frequency component of the signal in order to reconstruct the signal from the samples. If it is not, the spectrum of the sampled signal has overlaps which introduce errors into the reconstruction of the signal. These



overlaps are the cause of "aliasing" and other inconsistencies in the reconstructed signal. This result is formalized in the Sampling Theorem, or "theorem of uniform sampling":

> If a signal contains no frequency components for $|f| \geq W$, it is completely described by instantaneous sample values uniformly spaced in time with period $T_s \leq 1/2W$.

This rate of sampling, $f_s = 2W$, is known as the Nyquist Rate.

It is the absolute minimum sampling rate from which a signal can be reconstructed under ideal conditions.

Until now consideration has been given only to ideal sampling and reconstruction. In practice, the ideal is rarely, if ever, realized. Practical sampling differs from ideal sampling in three obvious respects:

(1) Practical sampling waves are not composed of delta functions, but have a finite width.

(2) Practical filters are not ideal.

(3) Real world signals normally sampled are not strictly bandlimited.

In practice, these differences are of small enough significance that for all normally encountered cases only the last is of real importance. Since most real signals are not strictly bandlimited, there will be some overlap in the sampled spectrum. If the spectral components outside some nominal value W are negligible, however, the signal can still be adequately described for most purposes by samples spaced $T_s \leq 1/2\,W$.

SELF TEST III

QUANTIZATION AND SAMPLING

1. Quantization of signals is necessary in order to

____________________________.

2. The number of quantization levels necessary to represent a signal is related to __________________________.

3. The number of samples necessary to represent a signal is determined by __________________________.

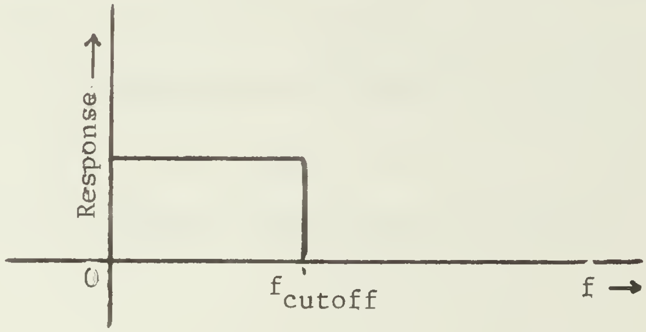
4. What is a "delta function"?

5. What is "aliasing"?

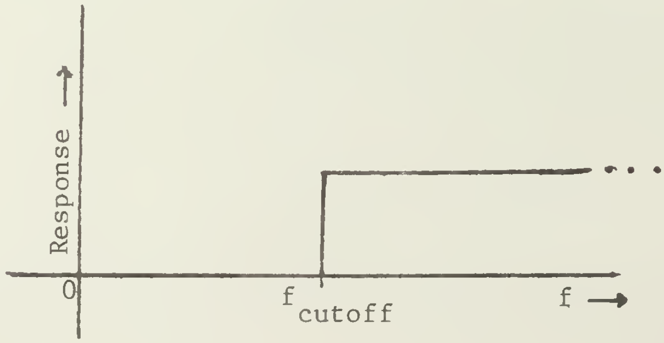
IV. FILTERS AND LINEAR SYSTEMS

The ocean is an extremely noisy environment in which to search for submarines. Noise from waves, wind, breaking surf, shipping, sonic mammals, fish and crustaceans, seismic activity, rain, etc., is present to some degree at all times. The problem then is how to go about sorting out these different sources in order to detect submarines. One way is to use filters to eliminate as much of the noise as possible.

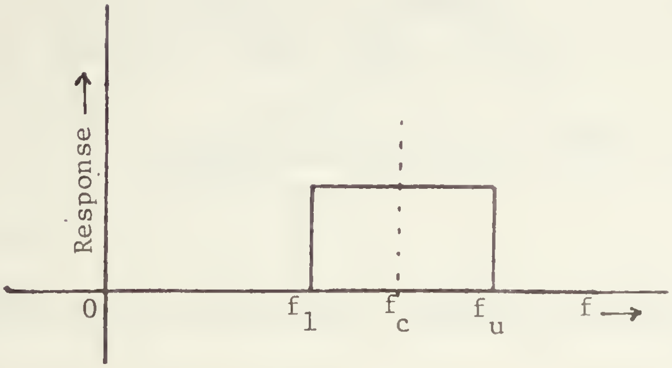
In general, there are several classes of filters relative to spectral response, e.g., low-pass, high-pass, and bandpass filters. Ideal low-pass filters allow frequencies below their cutoff frequency to pass and to screen out all above cutoff. Their spectral response is shown in the following diagram.



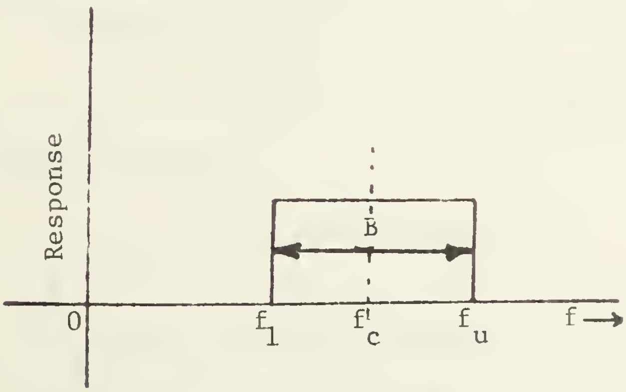
High-pass filters allow all frequencies above their cutoff to pass and to screen out all below.



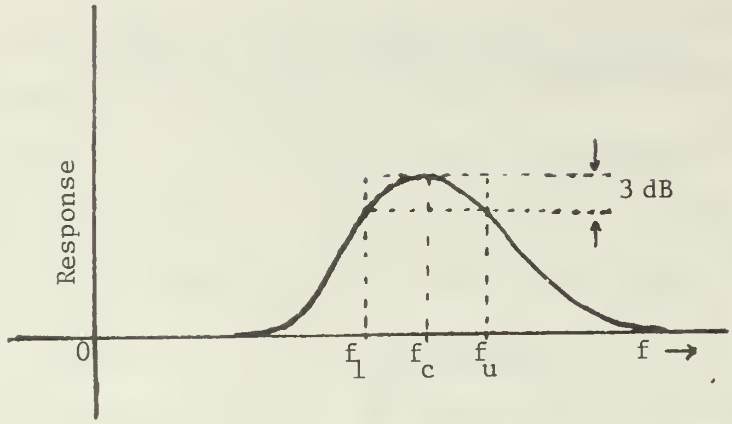
Band-pass filters allow all frequencies between their lower and upper cutoff frequencies to pass and to screen out all others.



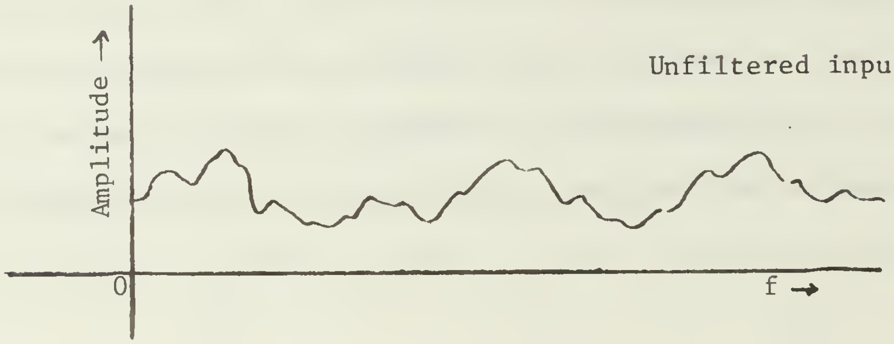
Filters are characterized by their upper and lower cut-off frequencies, or their center frequency and band-pass (or bandwidth), B. The bandwidth is defined as the passband width measured in positive frequency only, that is, the difference, $f_u - f_l = B$.

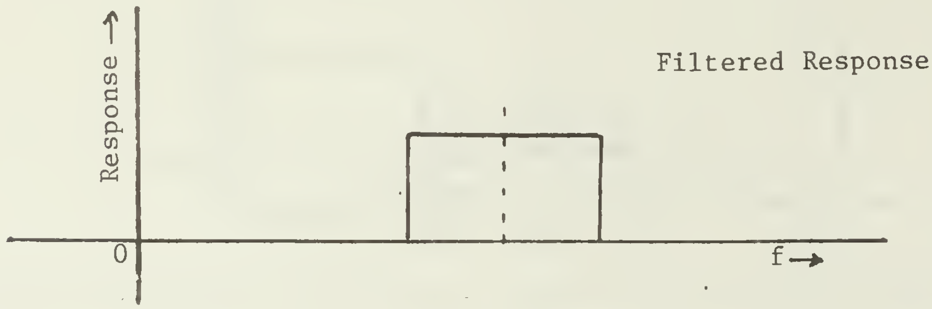


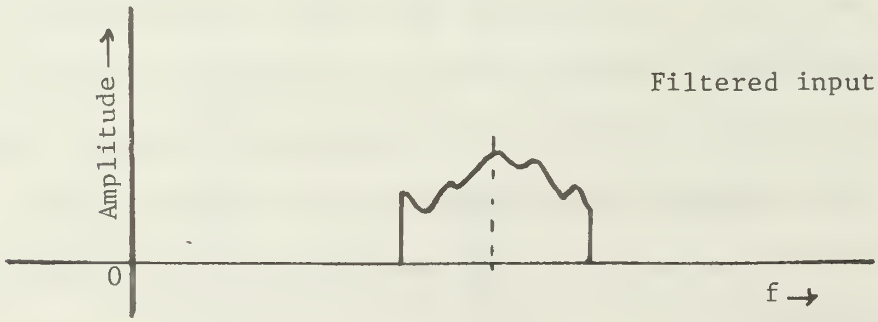
Of course no real filters have the sharp cutoff characteristics of the ideal filter, so it is customary to define the bandwidth of a real filter as the bandwidth between the points where the response has dropped to half the maximum. These points are known as the "3 dB down points", as shown on the following page.



A signal which is passed through a band-pass filter appears as follows:

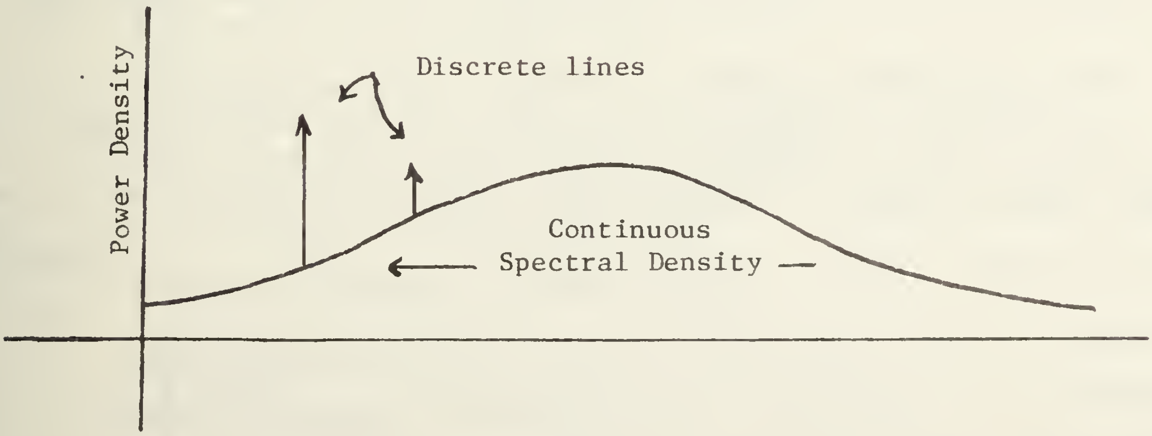






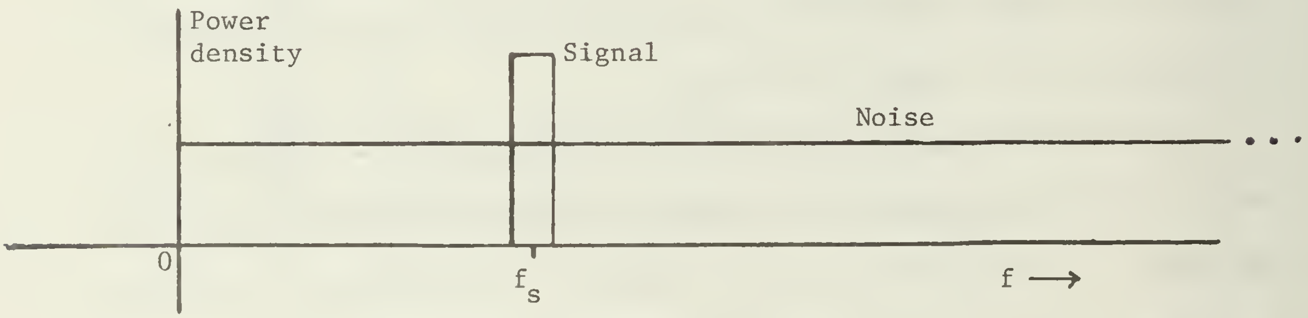
As indicated earlier, filters eliminate as much of the noise as possible. How does the filter actually improve detection capability? To answer this question, examination is made of the way filters help improve detection in a large class of detectors, known as "energy detectors". Discussion of the way these operate will be taken up in greater detail later. The general principle of operation is that the detector senses the total power (or energy) present in the input and compares this with a "threshold" value. If the input exceeds threshold, the detector indicates the presence of a signal and if the input does not exceed threshold, it indicates no signal present.

In order to understand how filtering helps to detect signals in noise, consider the "power spectral density" plot of an input. The power spectral density is the power in a signal that is carried by each frequency component.

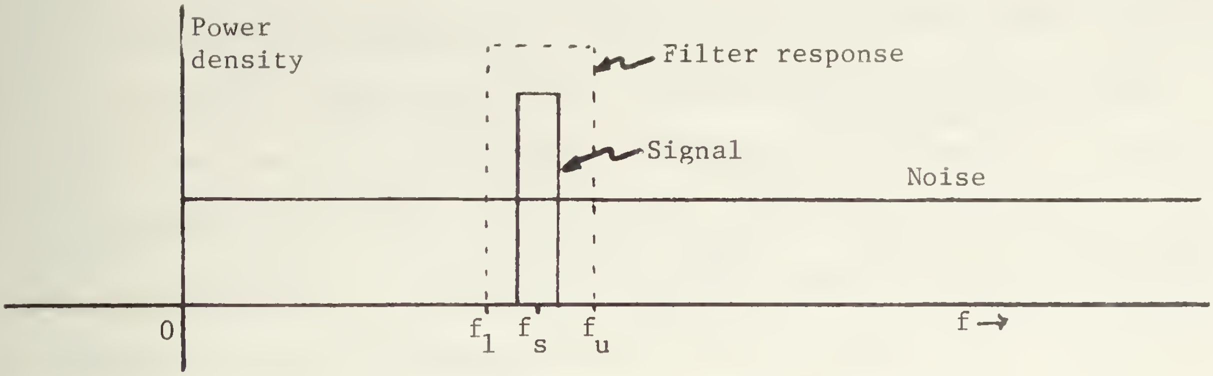


The total power in a signal then corresponds to the integral of the power spectral density over all frequencies. The power spectral density thus gives an indication of which

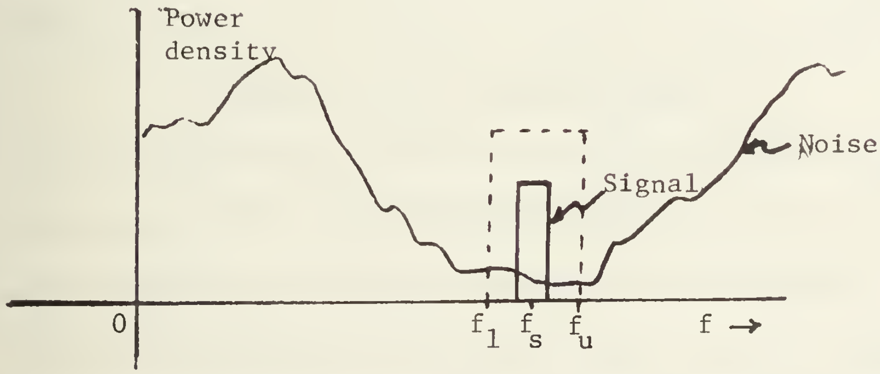
frequencies carry most of the power in a signal and the relative amount of power at a given frequency compared with that at others. The power in a given frequency band is then proportional to the area under the power spectral density curve between the lower and upper limits of the band. Comparison of the power present in a signal with the power of the background noise present indicates the detectability of the signal. To see the gains from filtering, consider a signal against a background of "white noise" (white noise is defined as having equal amounts of all spectral components). It is seen that



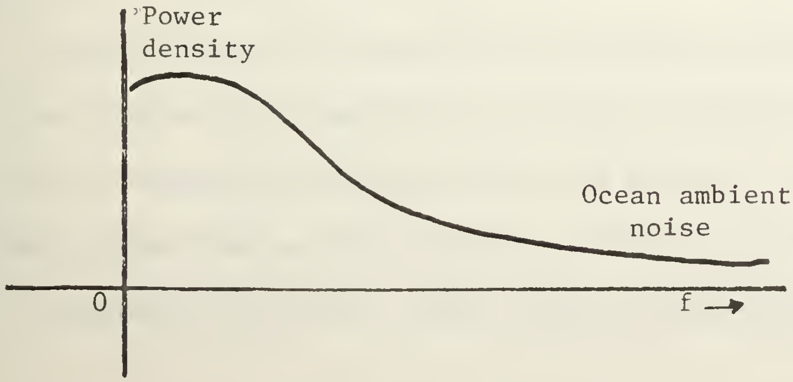
if the detector examines the entire frequency spectrum and senses the total power present to compare with a threshold value, the presence or absence of the signal does not affect the input significantly. On the other hand, if the input is filtered and only allows a narrow band of frequencies about the frequency of interest as the input to the detector, then the presence of a signal will substantially change the amount of power in the input, as shown in the diagram on the following page.



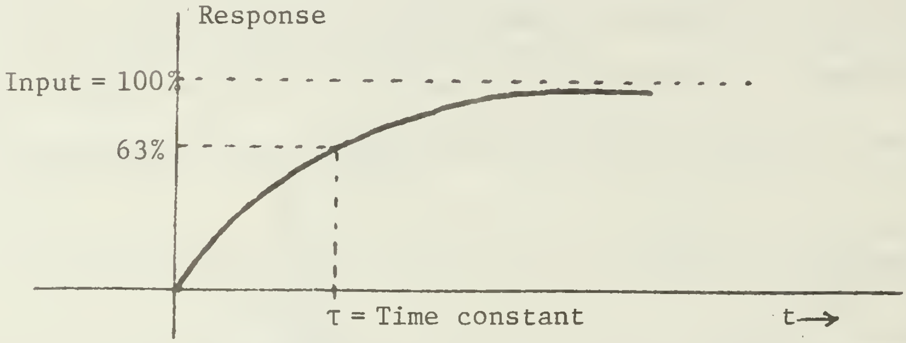
In effect, this raises the signal-to-noise ratio and thus allows detection of weaker signals than would be detected before filtering. In the case of non-white noise, the improvement in signal-to-noise (S/N) ratio can be even greater.



Ambient noise in the ocean is an example of non-white noise and careful filtering will improve detection probability immensely.



Another consideration in detection system design is the fact that many common filters in typical applications today are composed of reactive elements, such as LC/RLC circuits, crystals, or transducer elements, which have finite time constants associated with them. This means that these reactive elements require a finite time to "ring up", or establish resonance. Thus systems do not respond immediately to inputs. This must be considered in the system design. The response of these circuits has the following appearance:



The time required for the response to rise to 63% of the maximum is known as the time constant of the system. The narrower the resonance (the narrower the filter bandpass) the longer the time constant. Thus, if it is desired to resolve frequencies to $\pm B/2$ Hz, enough time must be allowed for the filter to respond (on the order of $1/B$ seconds) and this will be longer as B is made narrower.

Even modern digital filters require a finite time to respond to an input. This time is an inherent property of all signal processing systems. The general theory involved is called Linear Systems Theory. A "linear system" obeys the superposition theorem stated in Section II. Linear systems are completely described by their <u>transfer function</u>, H(f), which is defined

as the ratio of the output spectrum to the input spectrum.

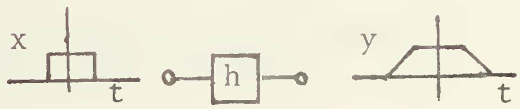
$$\boxed{H(f) = \frac{Y(f)}{X(f)}}$$

Their response to an arbitrary input is completely described by the impulse response h(t), which is the Fourier Transform of the transfer function.

$$\boxed{h(t) \leftrightarrow H(f)}$$

From Fourier theory, it can be shown that the output signal y(t) from a linear system is given by the <u>convolution</u> of the input signal x(t) and the impulse response.

$$\boxed{y(t) = x(t) * h(t)}$$



The <u>power</u> spectral density of the output, however, is given by the product of the input $G_x(f)$ and the <u>square</u> of the transfer function magnitude, often called the "frequency response".

$$\boxed{G_y(f) = |H(f)|^2 \, G_x(f)}$$

(See Page 67)

For example, a hi-fidelity audio amplifier is characterized by a flat frequency response from 20 Hz to 20,000 Hz, so that it does not filter out the musical content of a recording. However, the tone control on a car radio is provided to purposely filter out noise from "static".

Another consideration in the design of active ranging detection systems is the requirement to accommodate doppler in the returning echoes. The input filter must have a bandwidth wide enough to pass echoes that are doppler shifted by the maximum amount anticipated during operation of the system or the echoes will not be detected. If a filter of sufficient bandwidth to detect all echoes does not give a sufficient S/N ratio improvement, a bank of narrow band filters centered at various frequencies corresponding to different doppler shifts may be used alternatively.

SELF TEST IV

FILTERS AND LINEAR SYSTEMS

1. Name three classes of filters with respect to spectral response.

2. Explain how a filter can improve signal/noise ratio.

3. What is the "time constant" of a filter?

4. What is the "3 dB" bandwidth of a filter?

5. How is the bandwidth related to the response time?

6. What is the transfer function of a linear system?

7. How can the output of a linear system be calculated if the impulse response is known?

8. How can the impulse response be calculated if the transfer function is known?

9. What is meant by the "frequency response" of a filter (linear system)?

V. RANDOM SIGNALS, POWER SPECTRAL DENSITY, AND NOISE

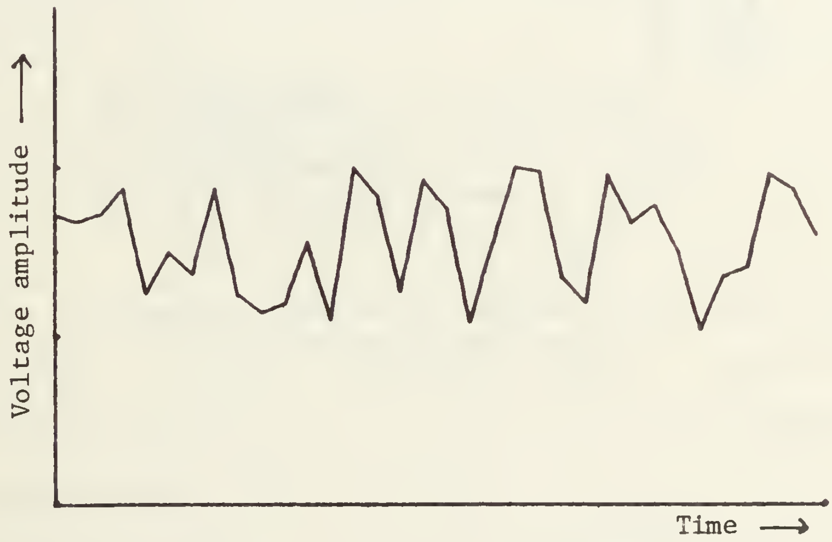
A. RANDOM SIGNALS

Heretofore the signals discussed have been explicitly described as "deterministic signals", such as a square wave, sinusoidal wave, pulse, etc. In describing a signal explicitly as some function of time, it is assumed that its amplitude and phase are known exactly for all time; i.e., past, present, and future. No real world signal meets these criteria for then the signal would convey no new information because it could be predicted exactly for any future time. Therefore, consideration must be given to random signals. Random signals of interest to the Antisubmarine Warfare specialist emanate, for example, from various machinery and flow noise. Noise in the ocean due to waves, wind, biological sources, seismic disturbances, etc., is a special case of random signal, and it is always present to some degree whether or not a "target" is present. Processing equipment must be able to distinguish between the random signal of interest, unwanted signals, and noise.

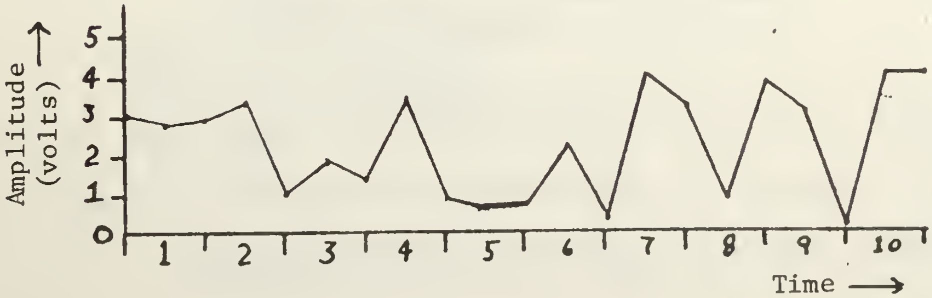
A random signal, unlike a deterministic signal, may <u>only</u> be described by its statistical characteristics. Its past history and its predicted future will be stated in terms of some <u>average</u> <u>value</u> and a given <u>probability</u> that the signal will be within certain limits at a specified time. This "given" probability" is represented by a probability density function (pdf).

What is a probability density function? A function is a descriptor, i.e., it usually describes something. Density is a measure of quantity containted within a specified space. Probability refers to the chance, or the percentage of time, an event occurs. Therefore, a probability density function describes the likelihood of some quantity being contained within a specified space.

How does the probability density function relate to a random signal? Assume a random signal, x(t), which has some randomly varying voltage amplitude over time.

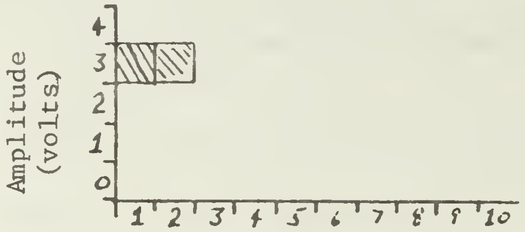


Plot the signal with a vertical scale representing voltage amplitude and the horizontal scale representing time.



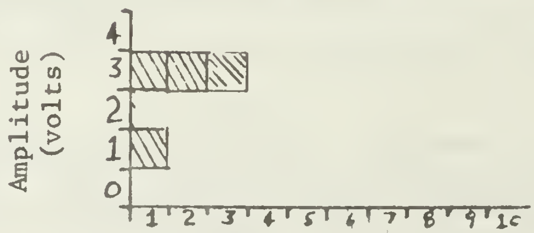
Suppose now that 1-volt represents any value between 0.5 and 1.49; 2-volt represents 1.5 to 2.49; 3-volt represents 2.5 to 3.49, etc. From the random signal, a plot is now made with the vertical scale still representing voltage amplitude but with the horizontal scale representing the number of times the signal is at that amplitude, or within that amplitude range.

During time interval 1, the signal was present twice in the 3-volt range and these two blocks are shaded in on the plot.



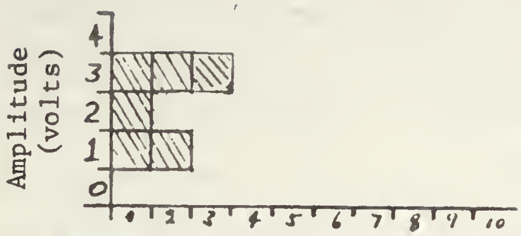
Number of times the signal is within the amplitude range

During time interval 2, the signal is present in the 1- and 3-volt ranges. These additional blocks are shown in the following diagram.



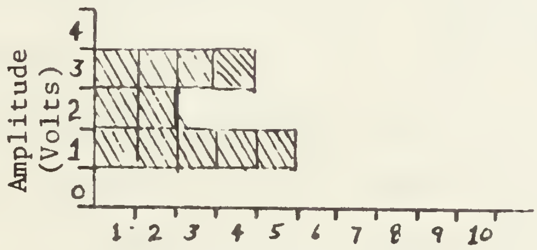
Number of times the signal is within the amplitude range

During time interval 3, the signal is present in the 1- and 2-volt ranges. These additional blocks are shaded in and are shown in the diagram on the following page.



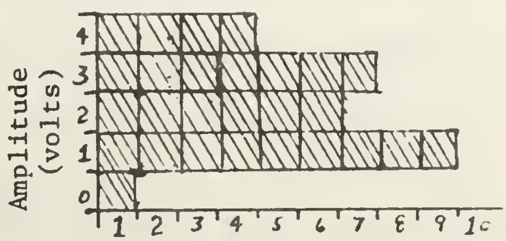
Number of times the signal is within
the amplitude range

The development should start to become apparent. Look now at time intervals 4, 5, and 6. The signal is present in the 3-volt range once, the 2-volt range once, and in the 1-volt range three times. Shade these in.



Number of times the signal
is within the amplitude range

To finish the plot, look at time intervals 7, 8, 9, and 10. The signal appears in the 4-volt range four times; in the 3-volt range three times; in the 2-volt range four times; in the 1-volt range four times; and in the 0-volt range once. Shade them in.



Number of times the signal
is within the amplitude range

The last plot on the previous page represents the amplitude ranges that the signal was in for the entire time it was observed. The total number of shaded blocks represents the number of times the signal is within the various amplitude zones during the various time intervals for the entire duration of the signal: the 27 blocks account for 100% of signal time. Each shaded block is regarded as a sample value. If the number of sample values is denoted by n, then n = 27 in this example. The percentage of time the signal is in a particular voltage range can be approximated by dividing the number of signal sample values in that voltage range by the total number of sample values, n.

0-volt range:	1/27 =	4%
1-volt range:	9/27 =	33%
2-volt range:	6/27 =	22%
3-volt range:	7/27 =	26%
4-volt range:	4/27 =	15%
	Total	100%

The percentage of time the signal spends at other ranges can also be calculated. Suppose, for example, that in a system there is a detector which detects only signals with amplitudes of 3 volts or more (the detector has a fixed "threshold" of 3 volts). To determine the percentage of time that the above random signal is detectable, it is necessary to determine the percentage of time it remains in the range of 3 or more volts. The 3-volt range covers from 2.5 to 3.49 volts. Take one-half the time the signal is in the 3-volt range and

add it to the entire time the signal is in the 4-volt range to approximate the total time the signal is at or above 3 volts.

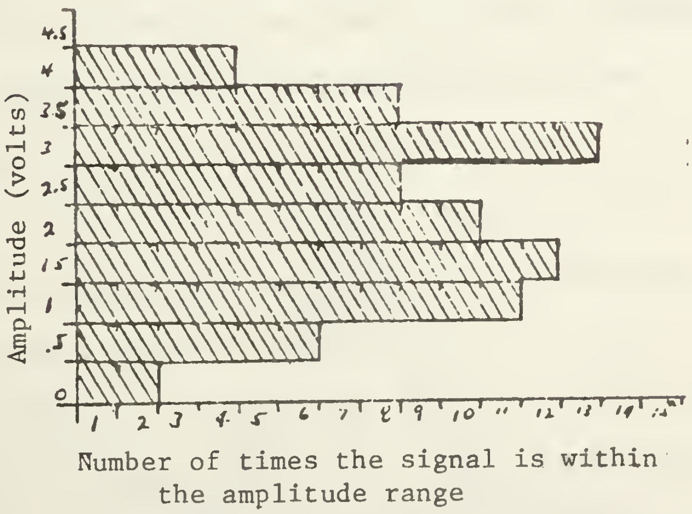
$$(1/2)\ (7) + 4 = 7.5$$

Divide by n to get percentage:

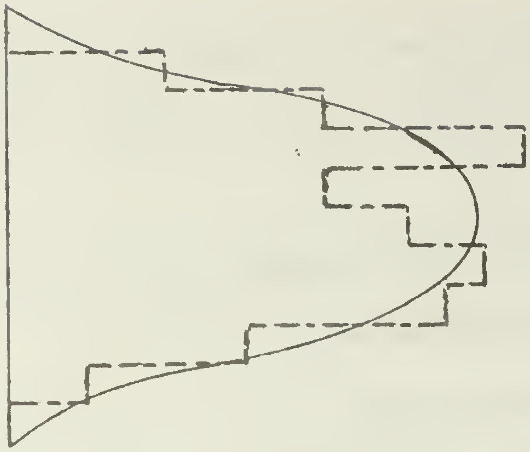
$$7.5/27 = 28\%$$

Therefore, the signal is detectable 28% of the time.

The accuracy in the determination of the percentage of time the signal is within any specified frequency range can be increased by making the voltage amplitude ranges smaller and shortening the duration of the time intervals. The following diagram shows sample values for the same signal, but with amplitude ranges of 1/2-volt and the time interval duration halved. It is constructed in the same manner as the previous plot.



As amplitude range and the time interval length are made smaller and smaller, the plot approaches a continuous curve as illustrated on the following page.



Continuous curve approximation

In most practical applications, a formula for the continuous curve can be determined and the corresponding function integrated over the amplitude range of interest to find the percentage of time the signal was within that range.

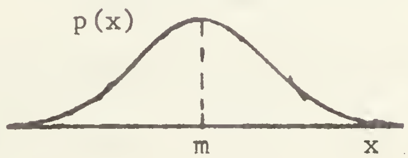
The probability density function of a signal represents the "probability" that a signal will be within an amplitude range. As the plot constructed above represents 100% of the time of the signal, so too, does the probability density function represent the total probability of the signal. It is scaled from 0 to 1, i.e., the area under the probability density function curve is equal to 1. If the detector in the example just mentioned detects, with certainty, any voltage amplitude of 3 volts or more, and the signal spends 28% of the time at 3 volts or more, then the probability that the signal will be detected in the 3-volt or more range is 0.28. The probability density is expressed by a function which may be integrated to find the probability that the signal is in any prescribed amplitude range. For probability density functions

used in many applied problems, the integration process can be avoided, since the results are readily available in standard tables.

The probability density function (pdf) most commonly used for electrical signals is the "gaussian" or "normal" pdf. It is the familiar bell-shaped curve described by the equation:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-m)^2}{2\sigma^2}} .$$

In the equation for p(x),

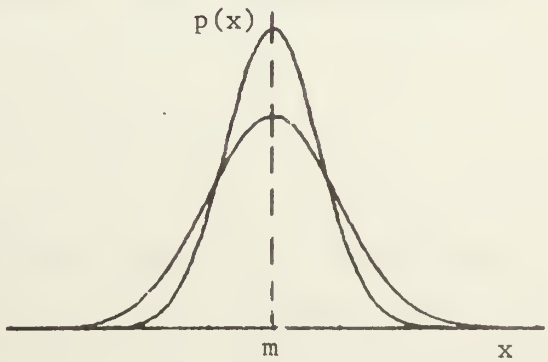


m = mean, or average, value,

x = value of signal x(t) at time t,

σ = standard deviation (σ is the Greek lower case sigma).

The standard deviation σ is a measure of the spread of the signal about its average value m. In the next figure is shown two gaussian curves with the same mean but different values of σ. The area under each curve is equal to 1.0.

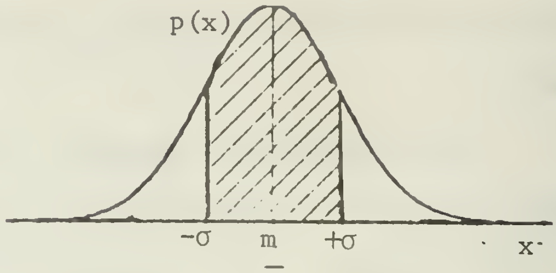


For the gaussian pdf, about 68% of the area under the curve lies between m − σ and m + σ, which means that the probability

P(·) that the random signal, x(t), assumes a value between $m - 1\sigma$ and $m + 1\sigma$ is 0.68.

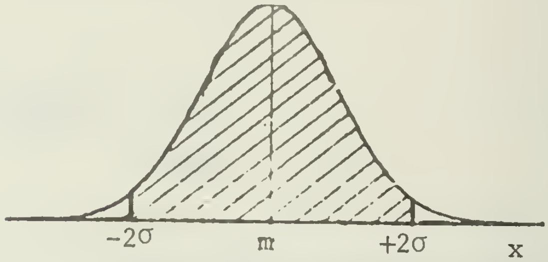
$m \pm 1\sigma$, $P(\cdot) = 0.68$

(Shaded area = 68% of total area)



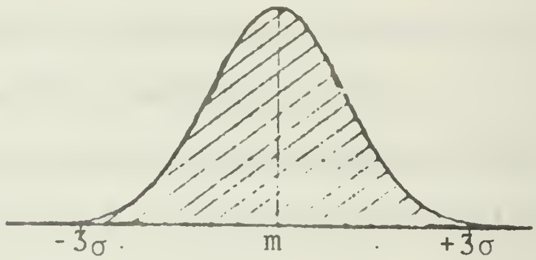
$m \pm 2\sigma$, $P(\cdot) = 0.95$

(Shaded area = 95% of total area)



$m \pm 3\sigma$, $P(\cdot) = 0.997$

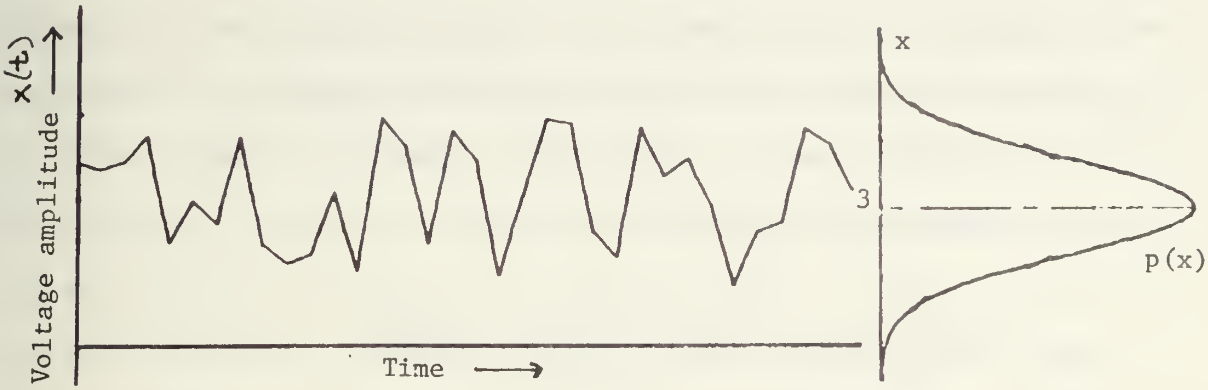
(Shaded area = 99.7% of total area)



In the illustrations, only 0.997 of the total probability has been accounted for. This is because there are no limits on a gaussian pdf; the range is unbounded. A signal may conceivably exist at any finite amplitude above or below the average value. The probability that the signal will occur outside the $m \pm 3\sigma$ range is only 0.003 ($1.0 - 0.997 = 0.003$). Limiting circuits are normally used early in the processing stage to prevent the rare occurrence of a large amplitude signal from damaging equipment. The important part of the signal, 0.997 of it, is still retained.

The gaussian pdf is centered around an average value. For an electrical signal, this is the average value of the

amplitude of the signal. Look at the signal and its pdf side by side with the pdf oriented to the signal amplitude in the following diagram:



The value chosen for m is the statistical average of the 27 sample values on the previous pages. Each of the sample values was 0, 1, 2, 3, or 4. The average, m, is equal to the sum of the individual sample values divided by the sample size, n.

$$m = \frac{0+1+1+1+1+1+1+1+1+1+2+2+2+2+2+2+2+2+3+\ldots+4\ldots}{2.7}$$

or, since many values are the same,

$$m = \frac{1(0) + 9(1) + 6(2) + 7(3) + 4(4)}{2.7} = 2.18$$

Algebraically expressed, this is

$$m = \frac{1}{n}\sum_{j=1}^{n} x_j$$

where

$\sum_{j=1}^{n}$ = the sum of all the samples

x_j = j = 1 to j = n, or $x_1 + x_2 + x_3 + \ldots + x_n$.

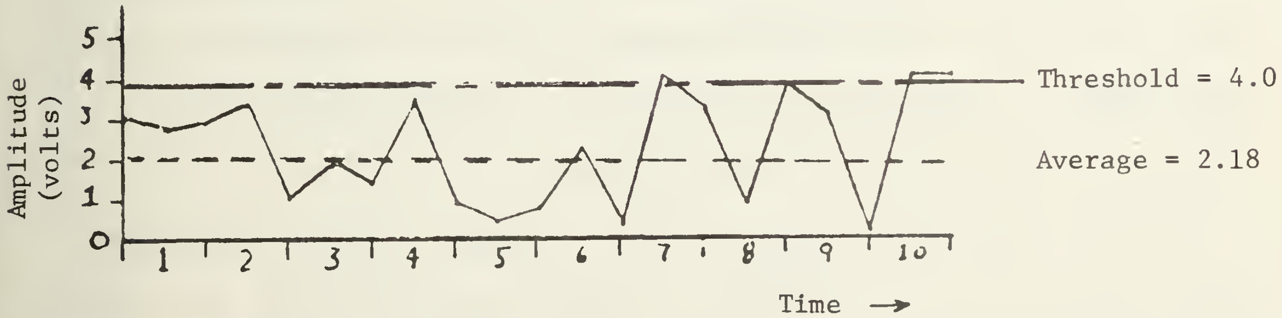
The expression given is equivalent to adding the sum of the products of the individual sample values and their probability of occurrence. Increasing the number of sample values gives a more accurate description of the signal. If m is calculated for a number of samples increasing without bound, the summation can be written as an integral in which the sample probability is replaced by the pdf p(x).

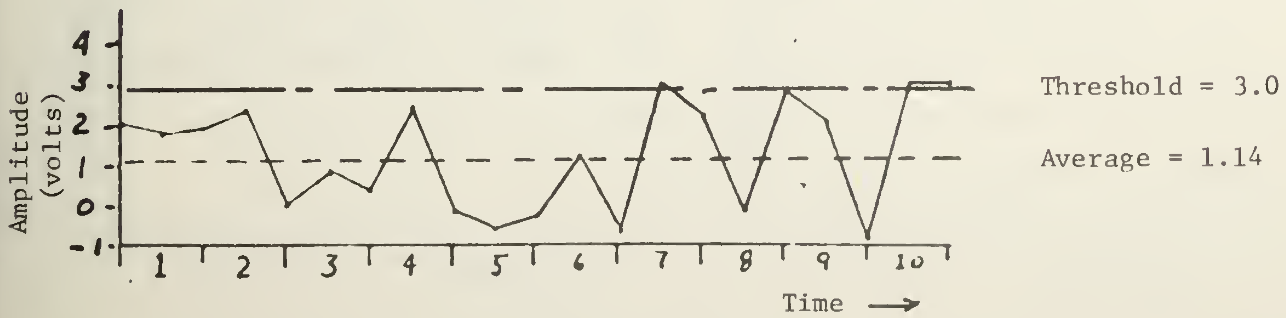
$$m = \sum_{j=1}^{\infty} \left(x_j\right) \cdot P\left(x_j\right) = \int_{-\infty}^{+\infty} x \cdot p(x)\, dx$$

It has been shown how statistical characteristics may describe a signal. The next problem is how to <u>obtain</u> signal characteristics that are <u>statistically</u> <u>valid</u> to identify a signal.

The validity of statistical methods in determining the outcome of a process depends on repetition of the process over and over again. With a random signal, x(t), this repetition may be accomplished by using an "ensemble" of random signal generators with each signal generated having the same broad statistical characteristics. If the individual outputs from this "ensemble" are measured <u>simultaneously</u> at some time, t, and averaged, the result would be an "ensemble average", indicated by $\overline{x(t)}$. This "ensemble average" is identical to, and has all the properties of, a statistical average. However, most signals of interest in typical systems originate from a single source. Therefore, only the time average of the signal, $\langle x(t) \rangle$, may be obtained.

The time average, $\langle x(t) \rangle$, and the ensemble average, $\overline{x(t)}$, are not always the same. For example, suppose the statistical characteristics of the signal from the random signal generators in the ensemble vary with time. Such a variation would not be reflected in measurements made at a fixed time, t_1. The ensemble average at time t_1 would be different from the ensemble average at time t_2. A random signal with these characteristics is called "non-stationary". If a detector is used to detect the random signal in the example above, what is the effect on the detector if the random signal is "non-stationary"? The average, m, was calculated as 2.18. Suppose that m varies with time and that over another time interval m is calculated to be 1.14. If the signal over the second time has the same characteristics as over the first, the effect on detection would be the same as raising the threshold to 4 volts over the first interval, as indicated in the following diagram.

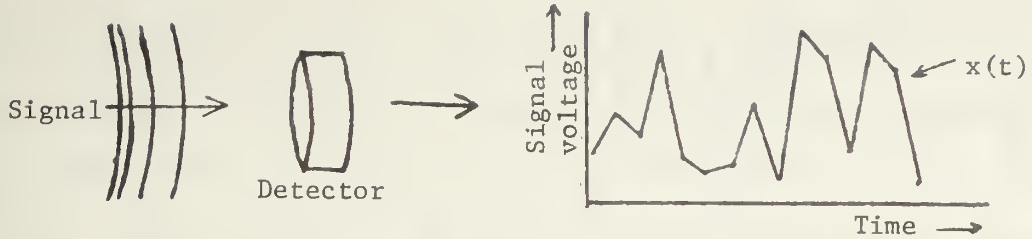




The probability is that $x = 3$ volts or more would become $2/27 = 74\%$ vice 28% for $m = 2.18$! If the average is unchanged and the standard deviation increased by a factor of three, the detector would be observing only 0.68 of the signal as opposed to 0.997. Therefore, if the random signal is not "stationary" (i.e., its statistical characteristics do not vary with time) or at least "stationary" for the processing time of the detector, the detector will be very inefficient. On the other hand, the signals might be stationary (such as a collection of batteries) yet be unequal to the ensemble average.

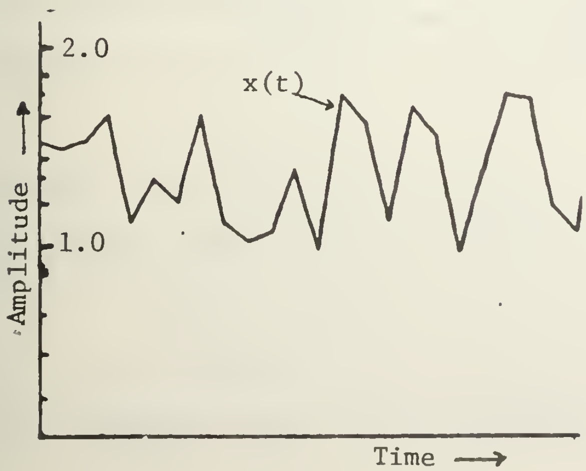
Fortunately, many random signals that must be processed may have time and ensemble averages which are identical, perhaps not for all time, but at least for the processing time required. Signals with their time and ensemble averages equal, $\langle x(t) \rangle = \overline{x(t)}$, are said to come from an "ergodic process", or to have the property of "ergodicity". An "ergodic process" is also stationary. Under the assumption that a signal from an ergodic process means the time and ensemble averages are equal, the signal is stationary and the time average will have all the properties of a statistical average.

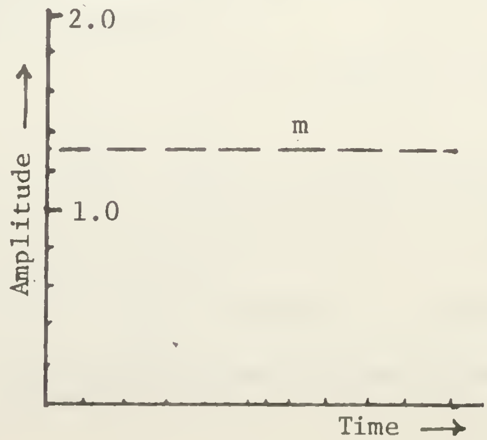
Having established that a time average will be statistically valid if it is from an ergodic process, how do we obtain the other statistical characteristics which describe a signal? All acoustic detectors convert the signal (and, unfortunately, the noise, which will be addressed later) into some time-varying voltage or current, $x(t)$.

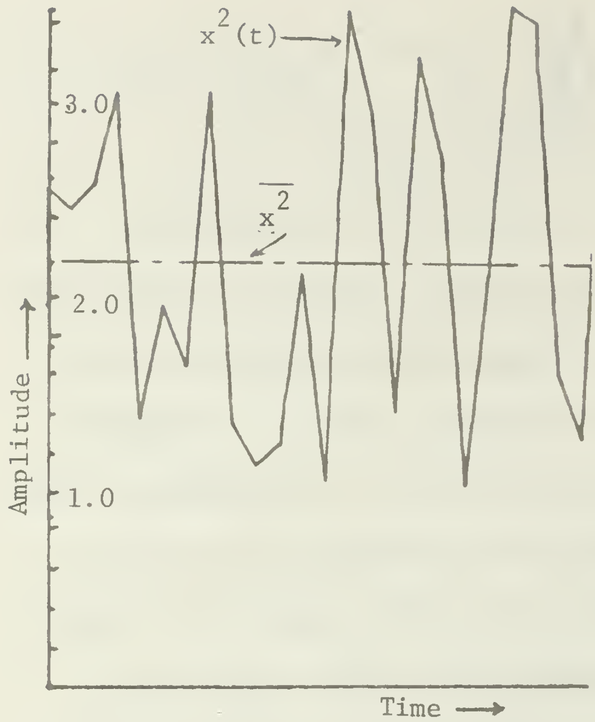


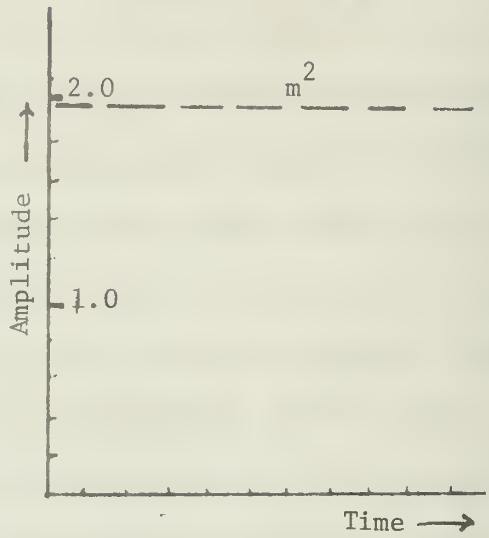
The time average of the signal, $\langle x(t) \rangle$, assuming ergodicity, is equal to the average value m. This average value is both the value about which the signal's density p(x) is centered and, as the time average, represents the "dc" component of the signal (direct current, also called the steady-state or the non-time-varying component). If m = 0, there is no "dc" component. If the signal x(t) is impressed across a one-ohm resistor, the instantaneous power dissipated is $x^2(t)$. The time average $\langle x^2(t) \rangle$ becomes the "second moment", which represents the total average power $\overline{x^2}$ of the electrical signal. The "second moment" differs from the "average value squared" m^2 which is the dc power, or the power in the non-time-varying component. m^2 is obtained by squaring <u>after</u> averaging, whereas $\overline{x^2}$ is obtained by squaring <u>before</u> averaging.

As an illustration of these values, consider a random signal x(t), diagrammed below and on the following page.

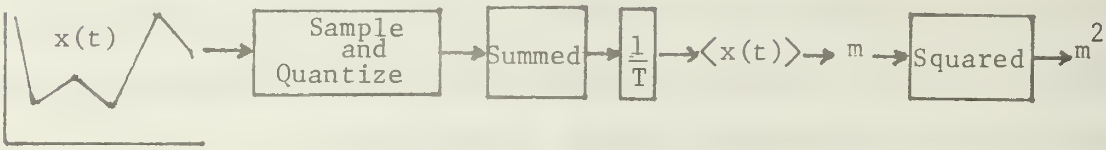


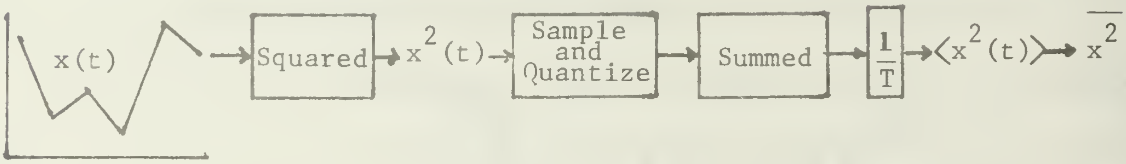






The following drawings illustrate computation of m^2 and $\overline{x^2}$.





There are two other measures which characterize any random signals in addition to the average value m, about which it is centered. A random signal is composed of a dc component and and "ac" component (alternating current, or the time varying

component of the signal). The total power $\overline{x^2}$ must be the sum of the dc power m^2 and the ac power σ^2. Although representing the ac power of the signal, σ^2 is the "variance" of the density $p(x)$. The variance is equal to the second moment minus the average squared, $\sigma^2 = \overline{x^2} - m^2$. For an electrical signal, the variance is the total average power minus the dc power. As the standard deviation σ represents the square root of the ac power, it is more commonly known as the root-mean square (rms) value. These definitions are summarized briefly:

Given an ergodic random signal $x(t)$:

1. the average (mean) value m is its dc component;
2. the average-squared m^2 is its dc power;
3. the second moment $\overline{x^2}$ is its total average power;
4. the variance $\sigma^2 = \overline{x^2} - m^2$ is its ac power;
5. the standard deviation σ is its rms value.

In "Fourier Transform Properties", the concept of correlation of a signal was discussed. It is considered again here because of its unique properties and because of its relationship to the "power spectral density" of a random signal. Those who feel that they have not grasped the concept of correlation should review that section before proceeding.

The autocorrelation function, $R(\tau)$, was defined previously as the time average $R(\tau) = \langle x(t) \cdot x(t+\tau) \rangle$. For an ergodic random signal, the autocorrelation function becomes:

$$R(\tau) = \langle x(t) \cdot x(t+\tau) \rangle = \overline{x(t) \cdot x(t+\tau)}$$

The value of $R(\tau)$ is a function of the time delay, τ, not time t.

$R(\tau)$ is a maximum at $\tau = 0$.

$$R(0) = \overline{x(t) \cdot x(t+0)} = \overline{x^2} = \text{total average power.}$$

As τ increases, $R(\tau)$ decreases, until at τ = infinity, $R(\tau)$ is a minimum:

$$R(\infty) = m^2 = \text{dc power;}$$

$$R(0) - R(\infty) = \sigma^2 = x^2 - m^2 = \text{ac power.}$$

The autocorrelation function is also a measure of the "time coherence" of the random signal. If τ is small, $x(t)$ and $x(t+\tau)$ will be close together in time. Therefore, the presence of one will have some effect on the presence of the other, a condition known as "statistical dependence". If the presence of $x(t)$ has no effect on $x(t+\tau)$, such as happens if τ is large and $x(t)$ is random and non-periodic, then $x(t)$ and $x(t+\tau)$ are said to be "statistically independent". If the absolute value of the autocorrelation function (minus the dc power) equals zero, then $x(t)$ and $x(t+\tau)$ are said to be "uncorrelated". For example, a sine wave is uncorrelated with a cosine wave but is statistically dependent. Statistical independence simplifies the calculation of the autocorrelation function. Whether or not two signals are correlated becomes important in the calculation of their average powers.

Assume that a signal has been formed by the addition of two separate signals: $z(t) = x(t) + y(t)$. The correlation function of $z(t)$ will be of the form,

$$(x+y)(x+y) = x^2 + xy + yx + y^2$$

and is equal to:

$$R_z(\tau) = \langle x(t) \cdot x(t+\tau)\rangle + \langle y(t) \cdot y(t+\tau)\rangle + \langle x(t) \cdot y(t+\tau)\rangle + \langle y(t) \cdot x(t+\tau)\rangle$$

$$R_z(\tau) = R_{xx}(\tau) + R_{yy}(\tau) + R_{xy}(\tau) + R_{yx}(\tau)$$

where the last two terms are the "cross correlation terms" of $x(t)$ with $y(t)$ and $y(t)$ with $x(t)$. As noted previously, the above correlation functions represent the average power in the signals, and can be written as:

$$P_z = P_{xx} + P_{yy} + P_{xy} + P_{yx}.$$

Let the part of the two signals $x(t)$ and $y(t)$ which is correlated over the time delay τ be called S. Then S is the sum of the coherent power S_{xx} in signal $x(t)$, the coherent power S_{yy} in signal $y(t)$, and their cross correlations S_{xy} and S_{yx}.

$$S_z = S_{xx} + S_{yy} + S_{xy} + S_{yx}.$$

If the part of the signals which is mutually uncorrelated over the time delay τ is called noise, then the total noise power N is the sum of only the noise power N_{xx} and N_{yy}. The cross correlations are equal to zero:

$$N_z = N_{xx} + N_{yy}; \quad (N_{xy} = 0; \quad N_{yx} = 0).$$

For a signal which is the sum of mutually uncorrelated signals, the correlation function is the sum of only the auto-correlations (the cross correlations equal zero) and the average power is the sum of only the average powers of the individual signals (the cross correlated powers equal zero).

B. POWER SPECTRAL DENSITY

Thus far only the time domain aspect of random signals has been addressed. The frequency domain is represented by the power spectral density, which was introduced in "Filters and Linear Systems", Section IV. The device used to transfer between the time and frequency domains is the Fourier Transform. The Weiner-Kinchine Theorem states that the power spectral density, $G(f)$, and the autocorrelation function $R(\tau)$ are Fourier Transforms of each other:

$$G(f) = \mathcal{F}\,[R(\tau)] = \int_{-\infty}^{+\infty} \langle x(t) \cdot x(t+\tau) \quad e^{-j\omega t}\, d\tau \, .$$

And from the Duality Theorem,

$$\text{if } R(\tau) \leftrightarrow G(f), \text{ then } G(\tau) \leftrightarrow R(f).$$

As can be seen above, the random signal may be described in either the time domain $R(\tau)$ or the frequency domain $G(f)$ with free interchange between the two.

Since the power spectral density is the Fourier Transform of the autocorrelation function, the Fourier Transform Theorems may be applied to the power spectral density. The systems designer uses the properties expressed by these theorems to design an efficient detector/signal processing system. The specialist who has knowledge of these properties can better understand how his system is designed to operate and, therefore, utilize it more effectively. The duality theorem was already utilized. The other major theorems applicalbe to power

spectral density will be explained along with examples of their usage.

1. Frequency Translation Theorem

If a signal $x(t)$ is bandlimited in $W < f_c$ and a new signal $y(t)$ is formed by multiplying $x(t)$ by $\cos(\omega_c t + \phi)$, $y(t) = x(t) \cos(\omega_c t + \tau)$, then the correlation function of $y(t)$ is equal to one-half the correlation function of $x(t) \cos \omega_c \tau$, that is, $R_y(\tau) = (\frac{1}{2}) R_x(\tau) \cos \omega_c \tau$. In addition, the power spectral density $G(f)$ is

$$G_y(f) = (\tfrac{1}{4}) G_x(f - f_c) + (\tfrac{1}{4}) G_x(f + f_c).$$

Multiplication in the time domain by $\cos \omega_c t$ becomes translated in the frequency domain as the power spectrum is shifted up and down in frequency by $\pm f_c$. The condition that the bandwidth W be less than f_c is necessary to ensure that $G_x(f - f_c)$ and $G_x(f + f_c)$ do not overlap, otherwise aliasing would result, as described in "Quantization and Sampling", Section III. Notice also that in multiplication by $\cos(\omega_c t + \phi)$ the phase factor ϕ is lost in translation to the power spectral density. Therefore, the original random waveform may not be reconstructed from knowledge of the power spectral density. However, by returning to the time domain via the inverse Fourier Transform, the autocorrelation function $R(\tau)$ with its significant statistical properties may be found.

Frequency translation is very useful in signal processing. Suppose that due to space, weight, or power limitations, it is impractical to locate the detector and the processor

together. The low frequency (0 to 1000 Hz) sound may be detected, translated to radio frequency (RF) such as used in acoustic listening sonobuoys and broadcast to the processor. In the same manner, one processor can handle many detectors if each detector's signal is translated to a separate frequency band within the RF range (Detector 1, frequency f_1 to f_2; Detector 2, frequency f_3 to f_4; etc., all within the RF band). The problem could be one of processing signals of many different frequencies. Rather than having many different processors each processing only one signal, it may be better to have one processor operating at a fixed frequency. All of the signals could be translated to the processor frequency for processing. If the processor is very elaborate and/or very expensive, this method becomes desirable.

2. <u>Integration Theorem</u>

$$\text{If } y(t) = \int_{-\infty}^{t} (t')\,dt' \text{ , then } G_y(f) = \frac{1}{(2\pi f)^2} G_x(f).$$

The multiplication factor $\frac{1}{(2\pi f)^2}$ indicates that the <u>low</u> frequency components of the power spectral density will be enhanced by integration. The ocean acts as an "integrating filter" in that it allows the low frequencies to propagate while rapidly attenuating the higher frequencies. This is one reason why, in an attempt to get long detection ranges, active sonars and passive processing systems have increasingly exploited the lower frequencies.

3. Differentiation Theorem

This theorem is offered without an example since it follows so naturally the integration theorem.

If $y(t) = d[x(t)]/dt$, then $G_y(f) = (2\pi f)^2 G_x(f)$. The multiplication factor $(2\pi f)^2$ indicates that the high frequency components of the power spectral density will be enhanced by differentiation.

C. NOISE

At the beginning of this section, noise was mentioned as a "special case of random signal which is always present to some degree whether or not a 'target' is present". In communications, noise is often defined as any electrical interference, or unwanted signal. Signal then, in addition to the characteristics previously used to describe it, has the quality of being wanted, sought after, emanating from a source of interest, and containing intelligence. In ASW, signal is always associated with "target". Noise, on the other hand, is anything which interferes with, or tends to mask, the signal. Often "one man's signal is another man's noise" and vice versa, such as passive sonar, acoustic countermeasures, etc. An acoustic detector converts both signal and noise into a time-varying electrical voltage or current. An examination of some of the characteristics of noise will help in understanding how systems may be designed to reduce it.

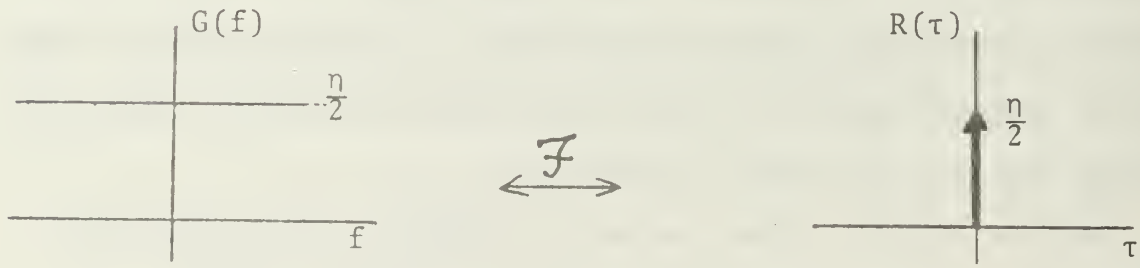
Another way to define noise is simply as that quantity observed in the absence of signal. Noise may come from a

variety of man-made or naturally occurring sources. Systems are designed to eliminate as much noise as possible, but some noise will inevitably remain. The thermal motion of electrons in the conducting media of the components of a system, for instance, is one unavoidable cause of electron noise. Thermal noise is interesting in that its power spectrum is constant over a wide range of frequencies. It is designated "white noise" by analogy to white light as all frequency components are present in equal amplitudes. The amplitude of thermal noise has been proven to have a "gaussian" pdf. It is therefore referred to as "gaussian white noise" ("gaussian", amplitude distribution; "white", frequency distribution). If "gaussian white noise" is played over a loudspeaker, it sounds dull and monotonous, somewhat like a waterfall. The subtleties of its random variations are hidden from the human observer.

The power spectrum of gaussian white noise, with zero mean, is $G(f) = \frac{\eta}{2}$, where η is power density in watts per Hertz. By Fourier Transformation, the autocorrelation function is

$$R(\tau) = \int_{-\infty}^{+\infty} \frac{\eta}{2} e^{j\omega\tau}\, df = \left(\frac{\eta}{2}\right)\delta(\tau)$$

where δ represents the "delta function" introduced previously.

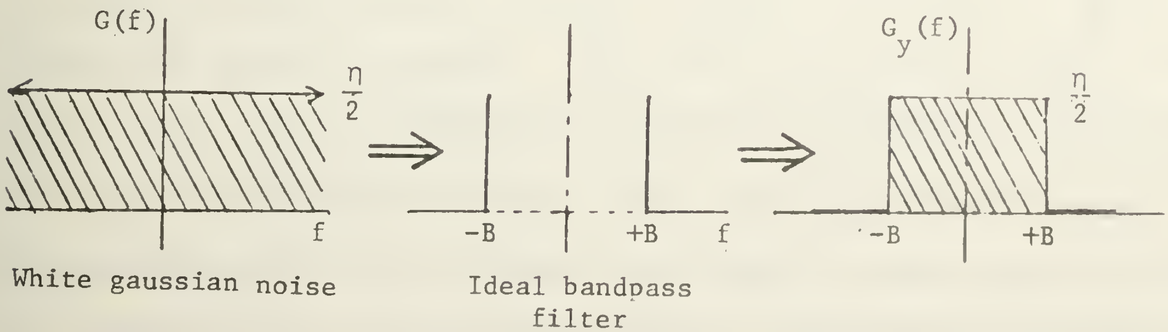


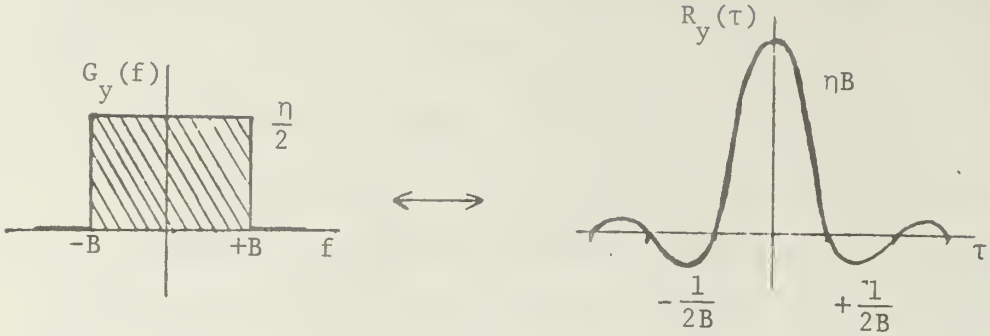
It is apparent from the $R(\tau)$ plot that for any time delay τ other than zero, the autocorrelation function is zero; therefore, any two different samples of a gaussian white noise signal are uncorrelated and statistically independent.

The signals of interest here are frequency limited, i.e., there is some finite frequency band within which the signal exists. Filtering in order to improve detection of that signal was discussed earlier. If gaussian white noise is filtered its frequency components are naturally those of the filter bandwidth but the amplitude distribution remains gaussian. The output power spectrum of white gaussian noise filtered by an ideal filter is a rectangular function,

$$G_N(f) = \frac{\eta}{2} \sqcap(f, 2B)$$

where $\sqcap$ indicates a rectangular function and B represents the filter bandpass. The autocorrelation function is the inverse Fourier Transform of the power spectral density. As indicated in "Fourier Transorm Properties", the Fourier Transform of a rectangular pulse is a "SINC function, $R_y(\tau) = \eta B \,\mathrm{SINC}\, 2B\tau$.





The figure above shows that filtering has produced the following results:

1. The power spectrum, though no longer white, is constant over the finite frequency range of the filter.

2. The output power is finite, $N = \eta B = \sigma_N^2$. Noise power varies linearly with bandwidth B.

3. The output signal is correlated over time intervals of about $\frac{1}{2B}$.

How is this knowledge used in signal processing?

Item 1. The noise in which a signal of interest must be detected is often not gaussian white noise. However, through judicious choice of a bandwidth, its power spectral density may be constant over that range. Otherwise, a pre-whitening filter may be used to make the noise power spectral density constant over the bandwidth. Therefore, after filtering it may be treated as gaussian white noise.

Item 2. The noise power after filtering is $\eta B = N = \sigma_N^2$. It is a direct, linear function of bandwidth B. By narrowing the bandwidth, the amount of noise power present is reduced. Narrowband (tonal) signals are enhanced in relation to the

noise power. This is crucial in a power, or energy detector.

Item 3. The bandwidth of the filter determines the minimum time delay τ to de-correlate noise. It must be greater than $\frac{1}{2B}$, $\left(\tau > \frac{1}{2B}\right)$, or the noise as well as the signal will be correlated. If the noise is uncorrelated, which it will be for $\tau > \frac{1}{2B}$, the cross-correlation power approaches zero, for example, at the output of a beam former for signals off-axis. (Refer to Section IX).

The bandwidth of the filter also determines the finite integration time to be used in determining the time average. The time average is indicated by $\langle\ \rangle$ (brackets) and is defined as:

$$\langle\ \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)\,dt$$

Replacing T by $\frac{1}{B}$, we obtain the output as $B \to 0$.

A review of the development of bandwidth versus integration time in the "Filters and Linear Systems", Section IV, may be helpful in understanding these relationships.

This section has been an attempt to describe the concepts of random signals, power spectral density, and noise. The intent has been to provide a basis for understanding their individual characteristics. A grasp of these characteristics is necessary to understand the methods of signal processing. Some common methods in current use, DELTIC, energy detectors, correlation detectors, and beam formers, will be presented in following sections.

SELF TEST V

RANDOM SIGNAL, POWER SPECTRAL DENSITY, AND NOISE

Given an ergodic random signal x(t), the electrical quantity corresponding to:

1. The average (mean) value is ________________.

2. The average squared is ____________________.

3. The second moment is ______________________.

4. The variance is ___________________.

5. The standard deviation is ____________________.

6. The autocorrelation

   $R(0) = ?$

   $R(\infty) = ?$

   $R(0) - R(\infty) = ?$

7. Filtered white noise has power given by

   $N = ?$

   where $\eta = ?$

   $B = ?$

8. Power spectral density is given by ________________.

VI. DELTIC AND FFT

In Section IV on Filters and Linear Systems, it was shown that a bandpass filter can serve two functions: (1) increase signal/noise ratio, and (2) resolve differences in frequency (e.g., doppler). These properties are utilized in the analysis of signals in order to detect and classify targets. However, the frequencies which the signal may contain are not usually known in advance, so it is necessary to cover a wide range of frequencies. In addition, they should be covered simultaneously in order not to miss an intermittent signal.

It was seen that one method of accomplishing this is the use of a bank of many filters arranged side-by-side in frequency. In the analysis of signals with discrete frequency components, it is an advantage to have the bandwidth of the filter as narrow as possible which enhances both functions of the filter, as stated above. However, in order to cover a wide range of frequencies in this manner, the number of filters required may become prohibitively large.

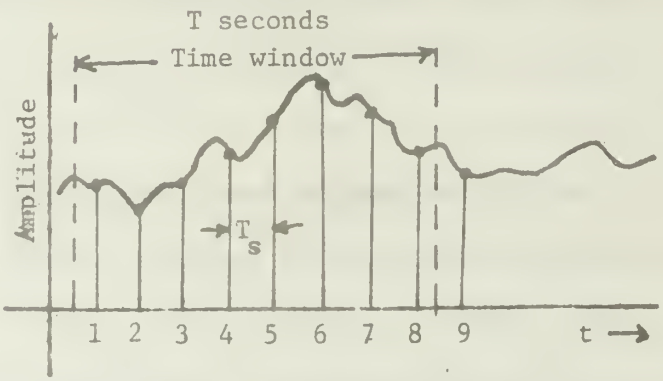
Another solution is to tune a single filter, like tuning a radio, in order to scan the band. In this case it is necessary to remain tuned to each frequency "bin" for a finite time.

It was shown that one must wait for longer periods of time as the frequency resolution requirements becomes stricter. For acoustic frequencies, however, the "ring up" time required by a single filter may be much too long to cover a band adequately. In order to circumvent this, several processing schemes have

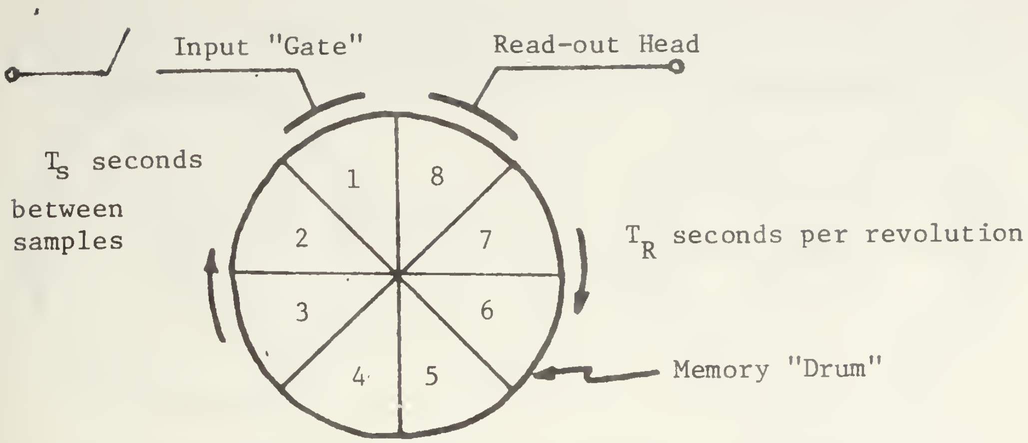
been developed which in effect give more filters for a given processing time.

DELTIC

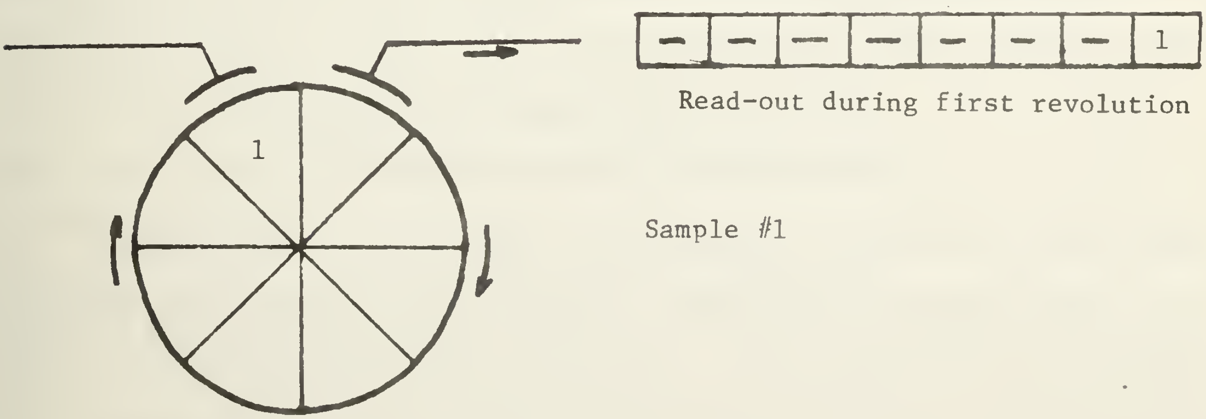
A scheme commonly used in acoustic spectrum analyzers is known as the DELTIC (DELay TIme Compressor). Investigation of the way in which this processor operates shows that processing the signal by a single filter requires a dwell time T, approximately equal to 1/B. The DELTIC takes the signal as received and records T seconds of data in a recirculating memory. It does this in a special sequence, however. To observe how this is processed, a simple signal is shown.



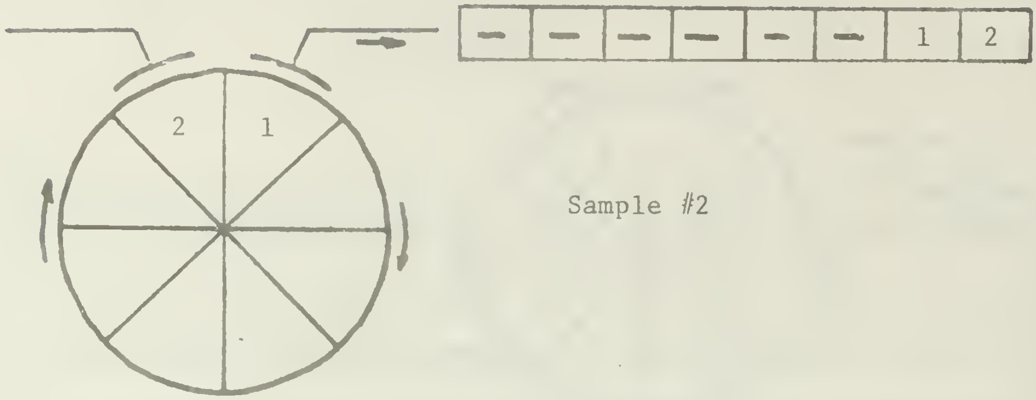
For simplicity in showing the operation, an 8-sample point "time window" will be used. In actual practice, 1000 or more sample points are commonly used. If the time between sample points is T_s, the time window observed will be $8T_s = T$. Therefore, if the DELTIC is not used, the best resolution obtained is 1/T. It would require T seconds to get this resolution. The physical layout of a DELTIC is shown in the diagram on the following page.



The "drum" is rotated at a rate T_R so that it makes one revolution plus one sample space during the time period T_s. In this case, with eight sample points, it makes 1.125 revolutions per T_s seconds. The read-out head reads out the sample values continuously as the drum rotates under it. The input gate injects a new sample point once each 1.125 revolutions in such a way that the sample point which has been on the drum the longest time is replaced. Starting with a blank drum, the process is shown as it occurs.

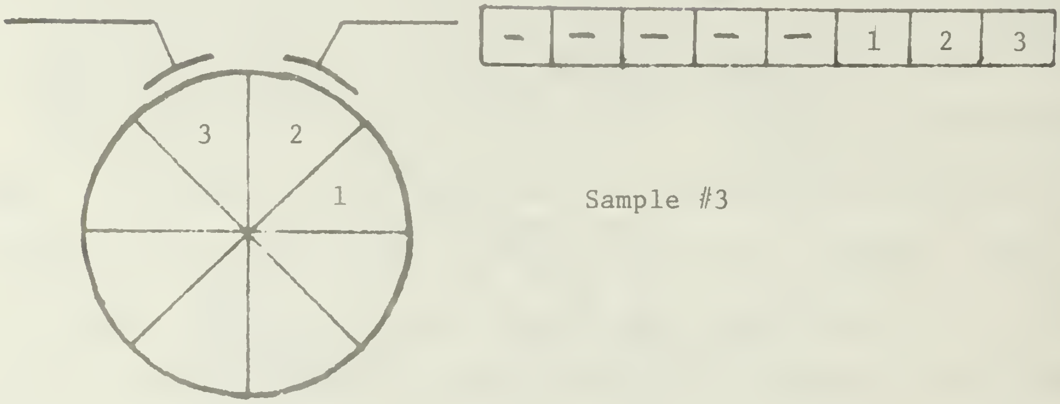


When the drum has made 1.125 revolutions (in T_s seconds), the input gate inserts the second sample point.



Sample #2

After 1.125 more revolutions, the third sample point is inserted.

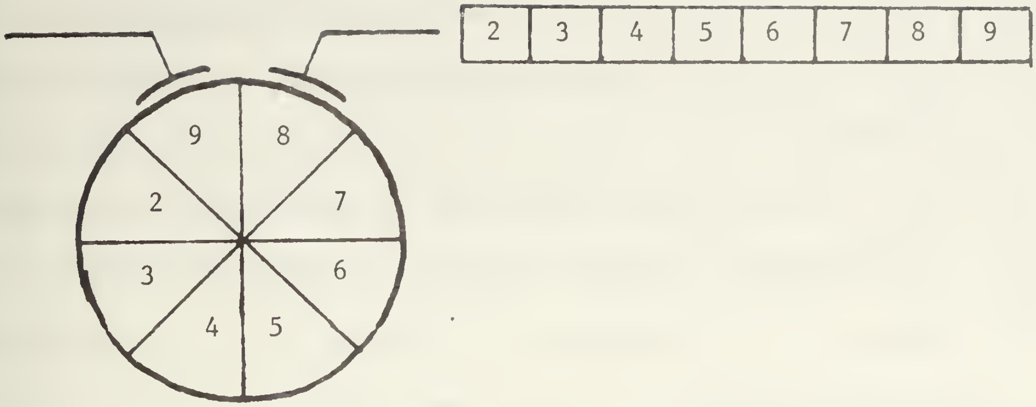


Sample #3

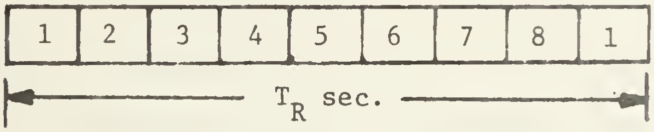
The remaining spots are filled similarly until all eight are filled. The output during one revolution of the drum is:

1	2	3	4	5	6	7	8

When the next sample point is inserted after 1.125 revolutions it displaces the first, as illustrated on the following page.

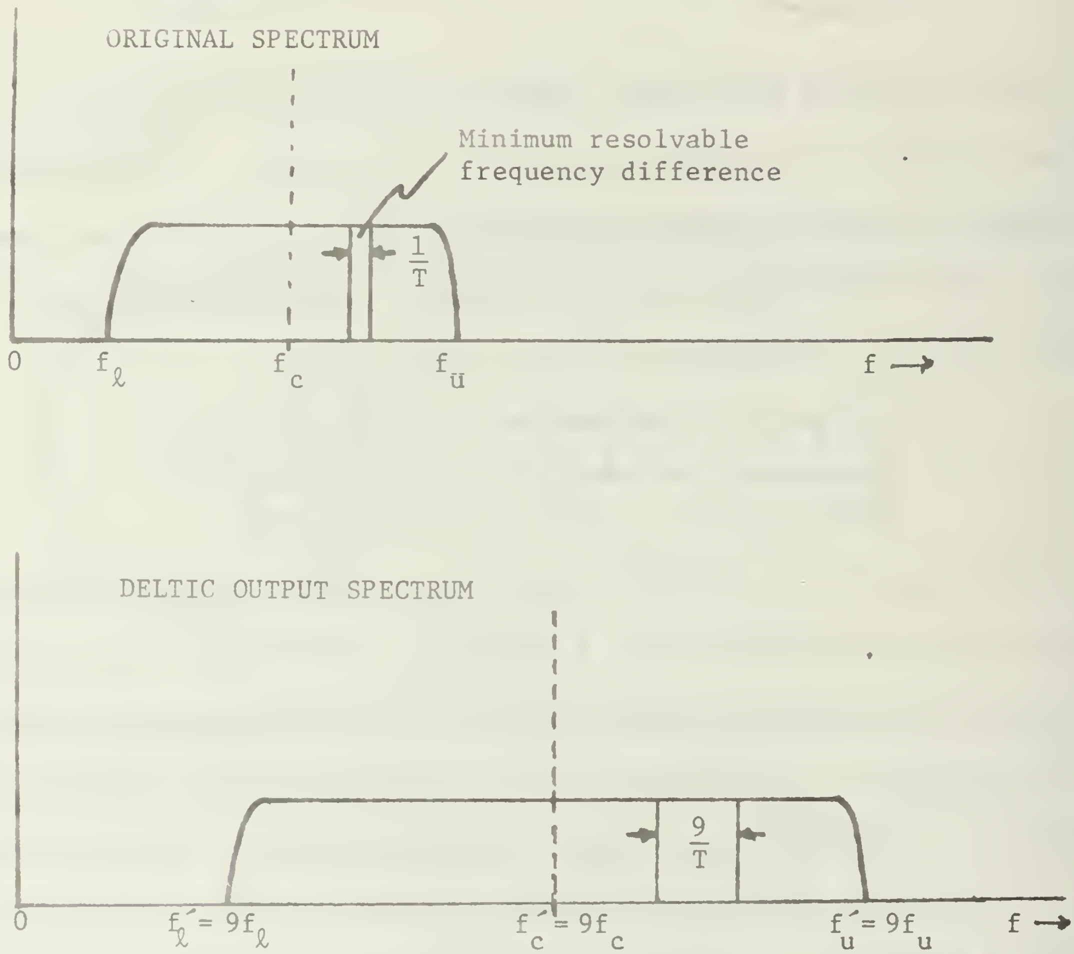


Once the drum is filled (which takes $8T_s$ seconds), a new value replaces the oldest one every T_s seconds. Looking at the output from the read-out head, it is seen that there are eight values read out each T_R seconds, plus one redundant sample which is insignificant for large T/T_s.



The output of the DELTIC each T_R seconds is thus the same as the inputs over the last T seconds. The DELTIC operation described above can be illustrated by the following analogy. If 90 seconds of conversation is recorded on tape and then the playback is speeded up so that the conversation takes place in only 10 seconds, the conversation has been compressed just as a DELTIC would do. The same information is in the 10 second playback as was in the 90 second conversation, but it only takes one-ninth the time to hear it. It is also noticed that this compressed conversation has been shifted in frequency by the compression and normal voices sound like Donald Duck. This illustrates the practical effects of the Fourier scale change theorem.

By using the DELTIC it is possible to observe the preceding T seconds' worth of signal during each interval of length T_R seconds. The frequency domain plots of the original signal and the output of the DELTIC show that the original spectrum spectrum has been shifted in frequency and stretched out.

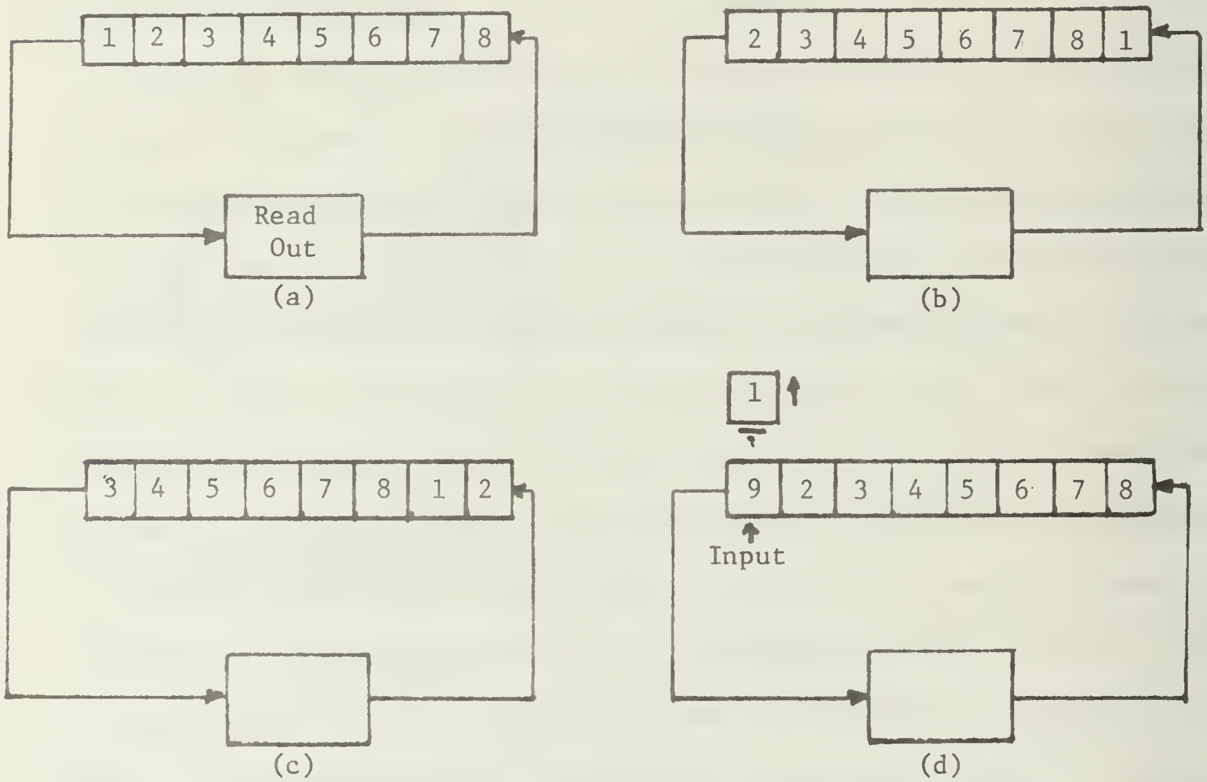


The Fourier scale change theorem shows that having compressed the time domain characteristics of the signal, x(t) becomes x(9t) and so the frequency domain characteristics of the signal X(f) becomes X(f/9) which has the effect of multiplying the frequency plot by nine.

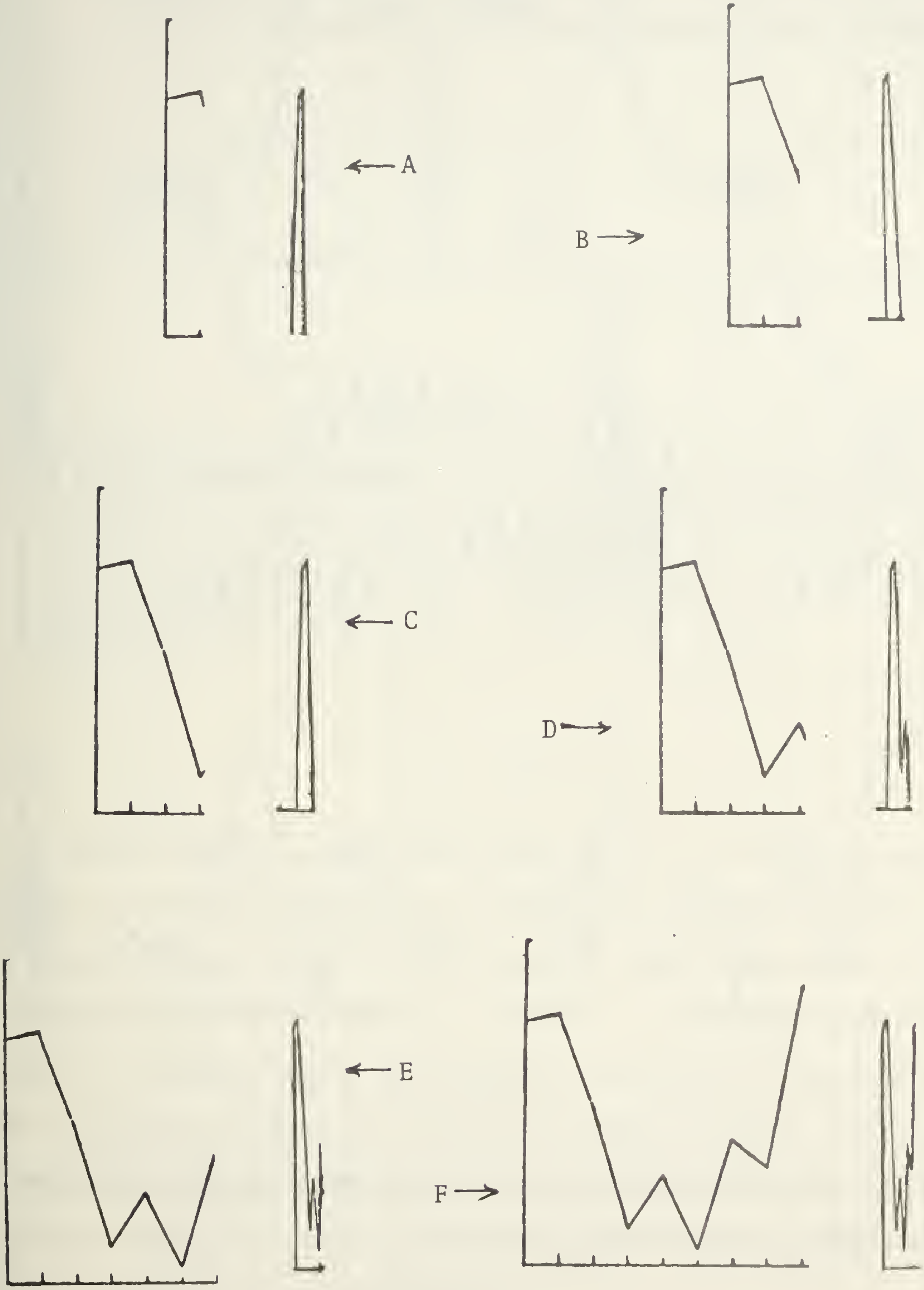
The "compression factor" is the ratio of the original sample time T to the compressed sample time T_R. Thus in this case it is the ratio T to T/9, or 9. Note that not only is the center frequency and bandwidth multiplied by this factor, but also the minimum resolvable frequency difference. Without compression, it would be integrated for T seconds to obtain a resolvable difference of 1/T. Upon compression, this becomes 9/T. At first glance, this does not seem to help, since the resolution has not been improved. However, the benefit in using the DELTIC is that this resolution can be obtained, not by integrating over T seconds, but over T/9 seconds once the DELTIC has been filled. Many spectrum analyzers "scan" the frequency band of interest by using a single band-pass filter of bandwidth equal to the minimum resolvable frequency difference of the processor and stepping its center frequency across the band one step each integration period. Thus, if the filter bandwidth is 1/9 the bandwidth of interest, without the DELTIC it would require nine integration periods (T) for one scan of the band. With the DELTIC, each integration period is T/9. Thus the entire band can be scanned with the same resolution in what was previously one integration time T. This is an advantage for a band of perhaps several hundred Hertz and a resolution of less than 1.0 Hz.

The DELTIC need not be designed for insertion of a new sample value every revolution, but only a new value every several complete revolutions. In this way the signal can be compressed even further by having the drum revolve more than once every T_s seconds.

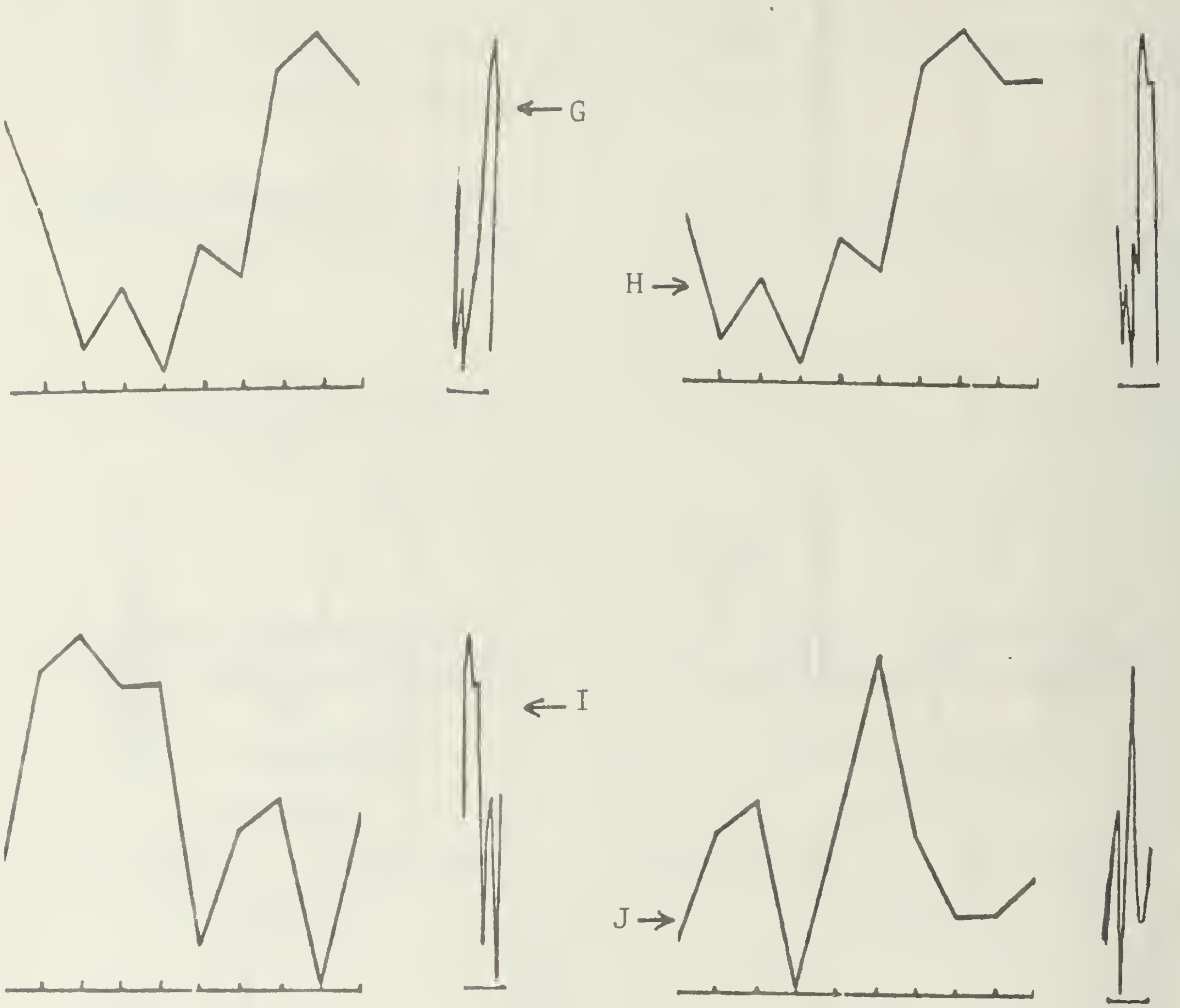
Instead of a drum, a delay line with feedback or a digital shift register may be used. The shift register is stepped past the read-out gate in such a way as to read the entire register each T_R seconds. The values are then recirculated. Each time a complete circuit is made a new value is input to replace the oldest and the process continues.



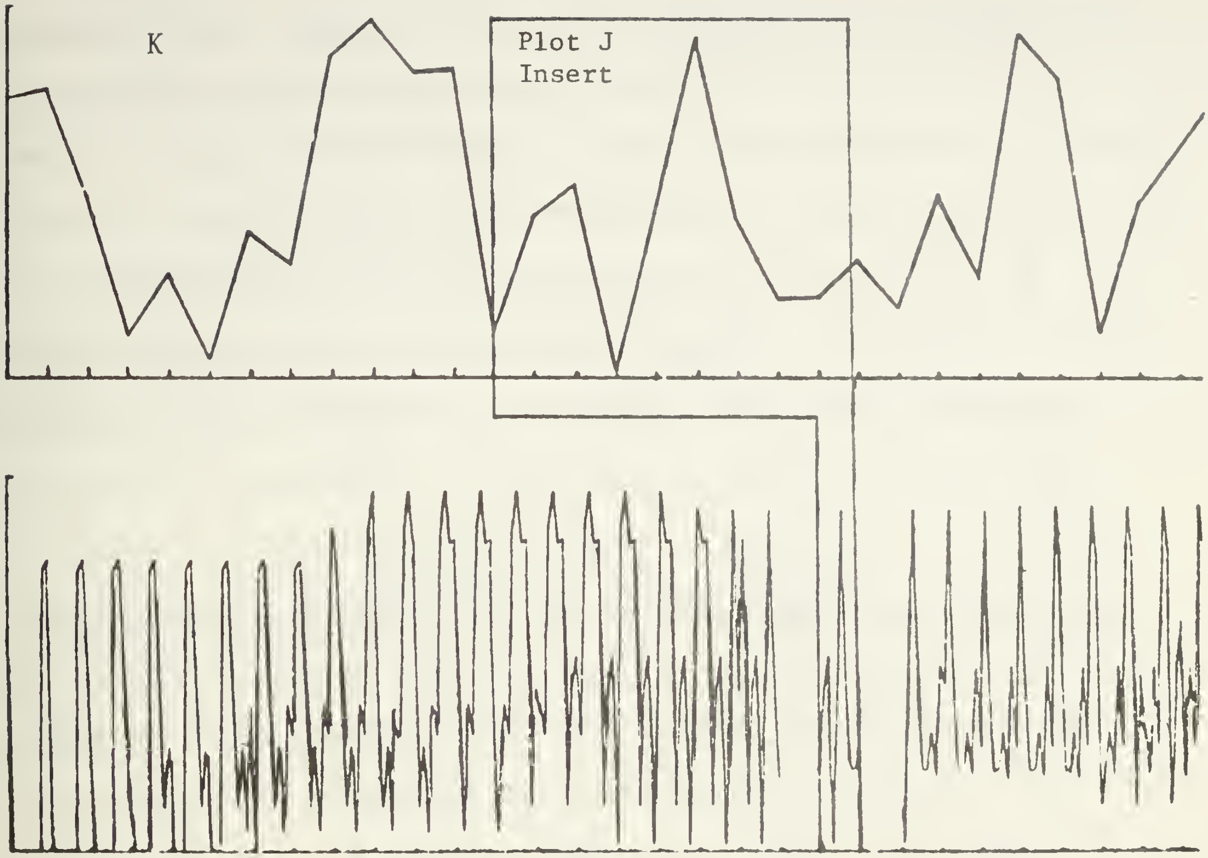
As the signal is "stepped" into the DELTIC, the output of the DELTIC represents the compressed signal. Plots A through F on the following page indicate the build-up of the DELTIC.


A
B
C
D
E
F

Plots G through J show the oldest information dropping out and only the most recent information being represented.



Plot K on the following page shows the complete time history of the signal and its DELTIC output.



FFT

Modern acoustic analyzers generally use a digital calculation of the Fourier series coefficients C_n from signal data observed during a time window T rather than processing it through a bank of filters or a DELTIC. As was discussed in Section III, Quantization and Sampling, in order to analyze a signal digitally it must be converted from a voltage ("analog" form) to a number. In fact, a continuously varying voltage becomes a list of numbers which must be stored in the computer memory for use in calculations.

The Fourier series requires data representing the signal for a finite period T, sampled at intervals T_s, taken

sufficiently close together to avoid aliasing. If the time T is not actually a complete cycle (or multiple) of the signal, the calculation can still be made but the result becomes an approximation to the spectrum rather than an exact analysis. If the signal is a pulse, the analysis approximates a Fourier Transform. The mathematical form of the calculation can be derived from the formula for the Fourier series coefficients, C_n:

$$C_n = \frac{1}{T}\int_0^T \left[x(t)\right]\left[e^{-j2\pi(n/T)t}\right]\left[dt\right]$$

Assuming that the signal has been sampled, x(t) becomes a list of sample values x_i taken at intervals T_s. The time t_i at which each sample was taken becomes iT_s and the total number of samples in period T becomes $N = T/T_s$. Substituting into a summation,

$$C_n = \frac{1}{T}\sum_{i=0}^{N-1} \left[x_i\right]\left[e^{-j2\pi(n/T)iT_s}\right]\left[T_s\right]$$

The formula becomes,

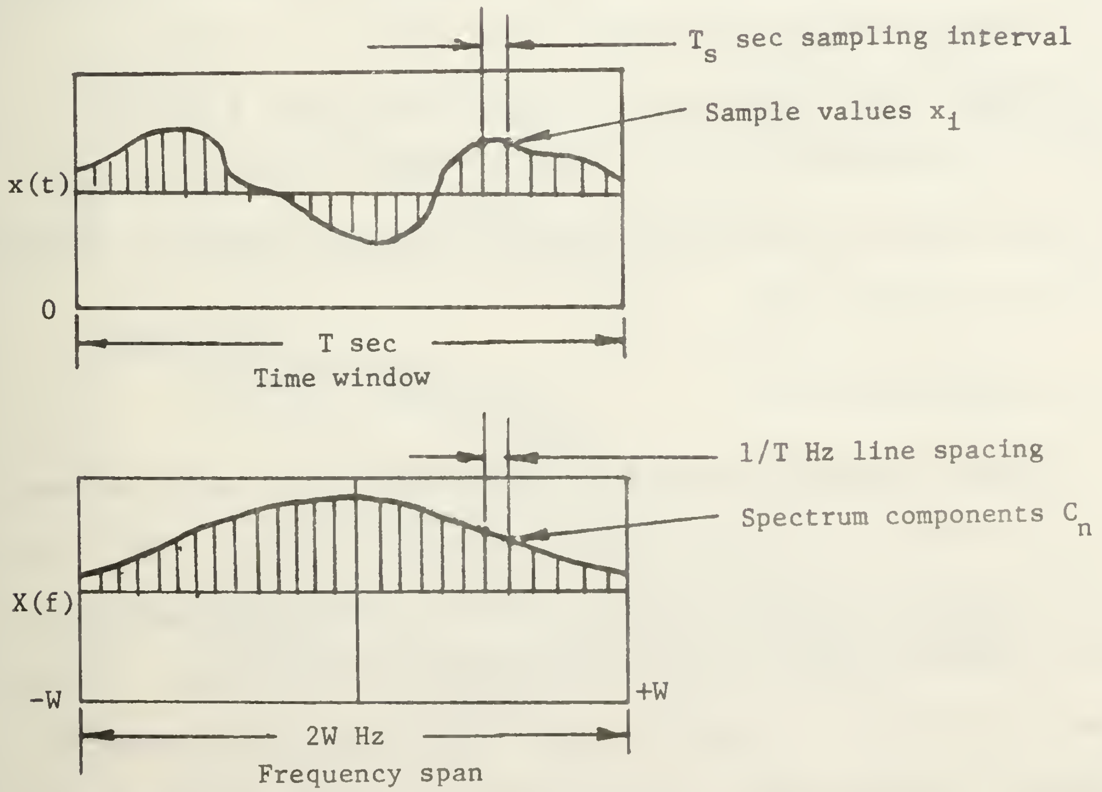
$$\boxed{C_n = \frac{1}{N}\sum_{i=0}^{N-1} x_i F^{ni}} \qquad \text{where } F = e^{-j2\pi/N}$$

Note at this point that the sampling must be accomplished at a rate for which $T_s = 1/2W$, where W is the highest frequency

component of the signal. Therefore, the number of samples processed for a time window T is

$$\boxed{N = 2\,TW}$$

Also, the highest order coefficient corresponds to W Hertz and since we are calculating 2-sided spectra, the total frequency span covers from -W to +W Hertz. Therefore, 2W Hertz divided by the separation of the Fourier components (1/T Hz) gives a total of 2 TW spectrum "lines". This number is called the <u>Time-Bandwidth Product</u> and is an important measure of the amount of data being processed and the memory required in the computer.



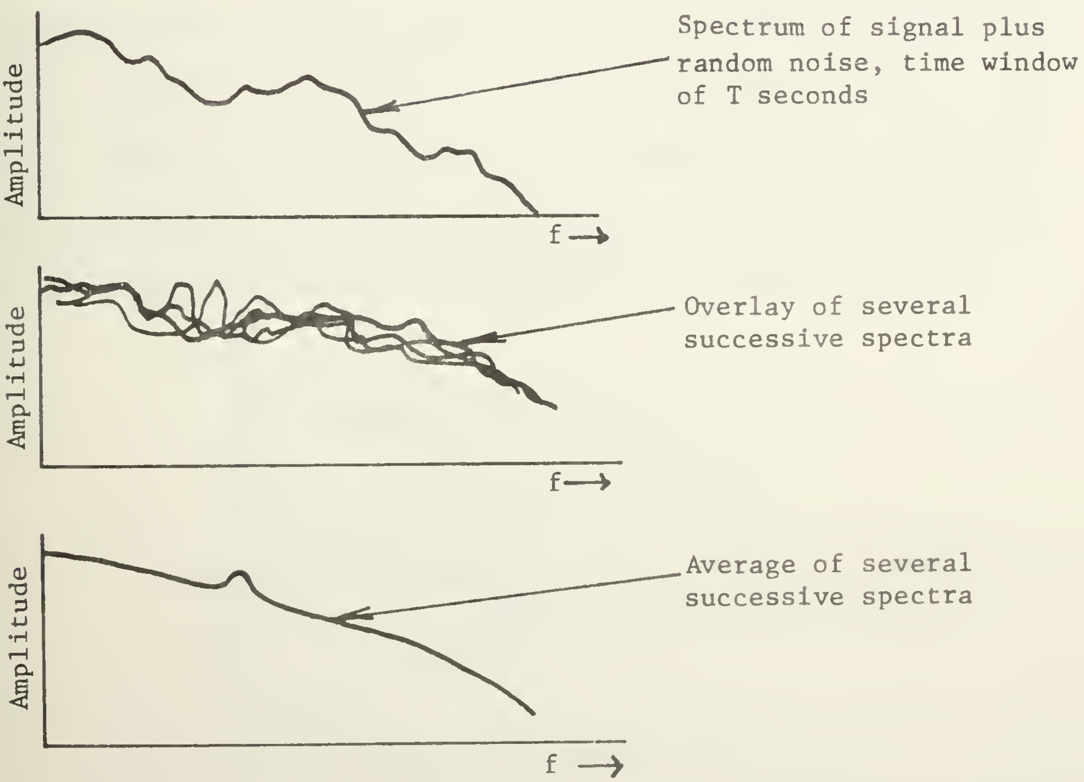
The summation formula requires N multiplications and additions for <u>each</u> spectral component so that a direct implementation in a computer would require N^2 operations to cover all N spectral lines. This calculation also has to be performed in T seconds in order not to miss any data. Even for acoustic frequencies, a digital computer would hardly be able to accomplish this calculation in "real time" if it were not for a clever algorithm known as the FFT (Fast Fourier Transform). This algorithm batches the data pairwise in a special way which allows the formula to be evaluated with only $N \cdot R$ operations, where N is equal to 2^R. The saving in time is on the order of N/R. For example, if $N = 2048 = 2^{11}$, the computation time is reduced by a factor $2048/11 = 186$.

Note that the FFT requires a time window T seconds long to provide frequency resolution of 1/T Hz, just as a filtér requires a dwell time T for the same resolution and a DELTIC requires T seconds to fill the "drum". This illustrates the fundamental nature of the relationship between frequency resolution and processing time.

SPECTRUM AVERAGING

If a random signal or noise is processed by a spectrum analyzer, the output spectrum is itself random as well. The presence of a discrete signal spectrum line in the noise will be masked by random peaks in the noise spectrum. However, if the spectrum is measured repeatedly, the signal line will consistently appear, whereas the noise peaks will fluctuate from measurement to measurement. Detection of discrete lines can be

enhanced by spectrum averaging, sometimes called "line integration". This averaging takes place visually by the operator if the spectra are recorded on a paper display, called a "spectrogram" or "gram" for short. The paper is marked by a stylus which travels across the display in proportion to the frequency, marking the paper in shades of gray proportional to the spectral density. As the paper advances, consistent discrete frequency components leave a linear trace, while the random noise background leaves a salt-and-pepper speckled pattern. The eye can visually "integrate" the spectrum and detect very faint lines. The spectra can also be averaged digitally, which is called ALI (Automatic Line Integration). Spectrum averaging further enhances weak signals but again processing gain must be paid for with time.



SELF TEST VI

DELTIC AND FFT

1. DELTIC is an acronym for ________________________________.

2. A DELTIC requires a ____________________________memory.

3. The advantage of a DELTIC when combined with a tunable filter is ____.

4. FFT is an algorithm which calculates ___________________.

5. The advantage of the FFT over direct calculation is

_______________________________.

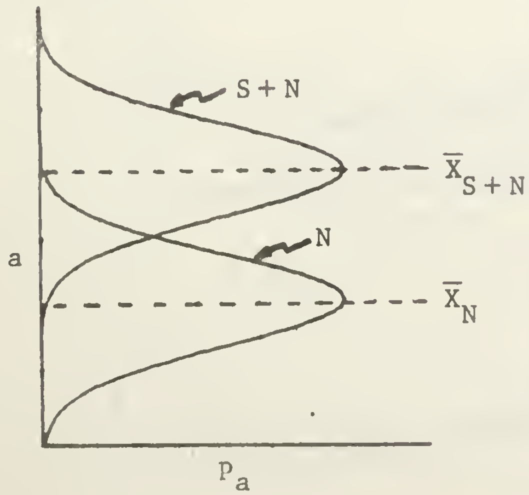
6. The Time-Bandwidth Product is a measure of ____________.

7. Describe how spectrum Averaging enhances detection of weak discrete signal lines.

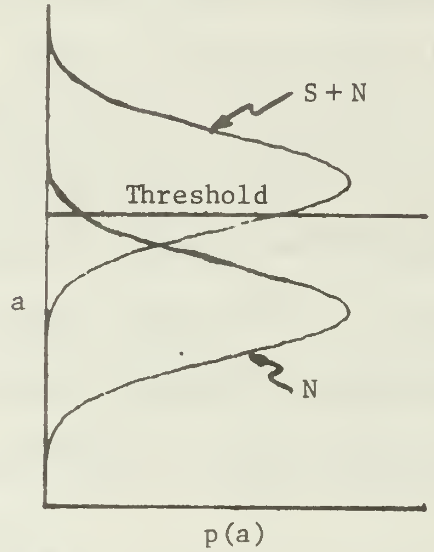
VII. ENERGY DETECTION

In the discussion of random signals, it was noted that in order to apply statistical analysis to signal processing, the signals must be "ergodic". One property of ergodic signals is statistical stationarity, that is, the statistical properties of the signal are constant in time. In reality, the signals that must be processed are not strictly stationary. The sea noise varies with the time of day, changes in weather, etc., but the variation is over a long time period. This fact makes it possible to apply the statistical analysis approach to signal processing as long as the processing time is short enough that the variation in signal properties is negligible.

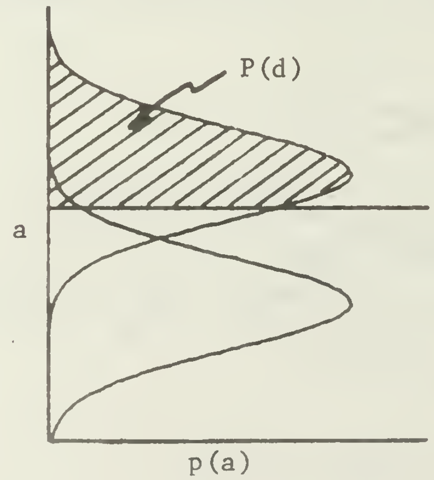
Consideration must be given to how to apply statistical methods to detection problems. For simplicity, the case of Gaussian noise is cited, where the probability distributions of amplitude for noise alone and noise-plus-signal are Gaussian with equal variance σ^2 but different means (average values).



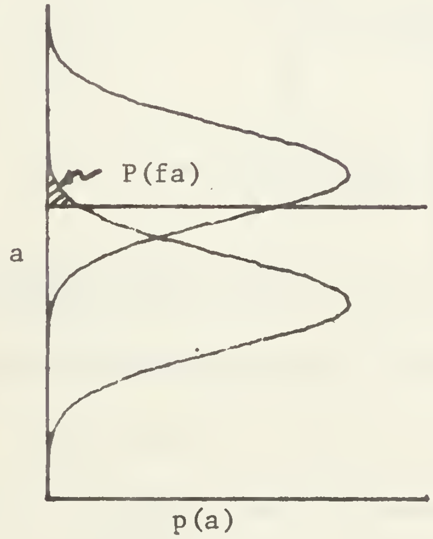
If there is now a processor which compares the amplitude of the input with a "threshold" level set into it and indicates "signal present" if the input is above threshold, "signal absent" if the input is below threshold, then statistical theory can be applied to determine the operating characteristics of the processor.



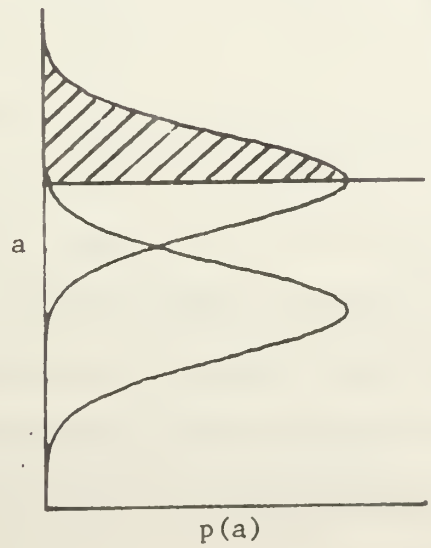
The area under the S + N curve above the threshold is proportional to the probability that the amplitude of the input will be above the threshold when the signal is present. Thus, this area is proportional to the signal detection probability, P(d).



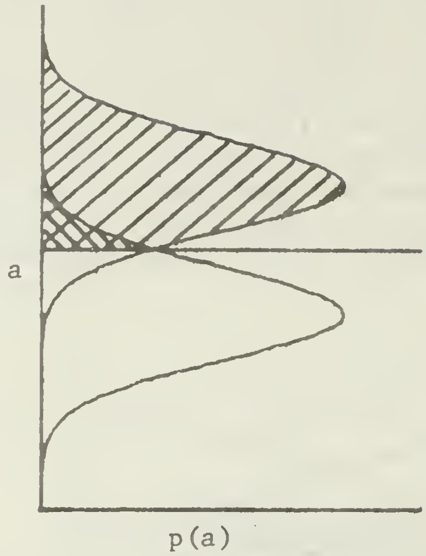
The area under the N curve and above the threshold is proportional to the probability that the amplitude of the input will be above the threshold, even when the signal is absent, because of the background noise. Thus, this area is proportional to the probability of false alarm, P(fa).



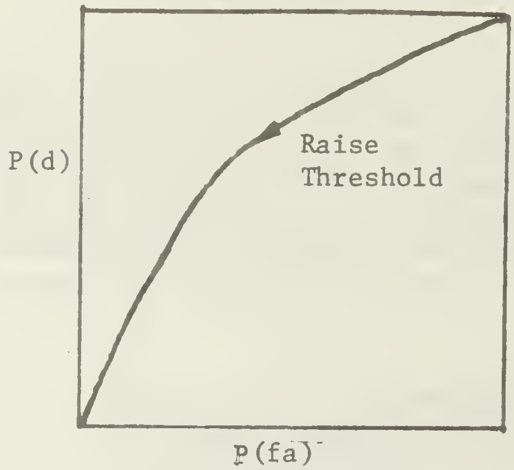
If the threshold is varied, the probabilities change. If the threshold is raised, the probability of detection decreases, but so does the probability of false alarm.



If the threshold is lowered, P(d) increases, but so does P(fa).

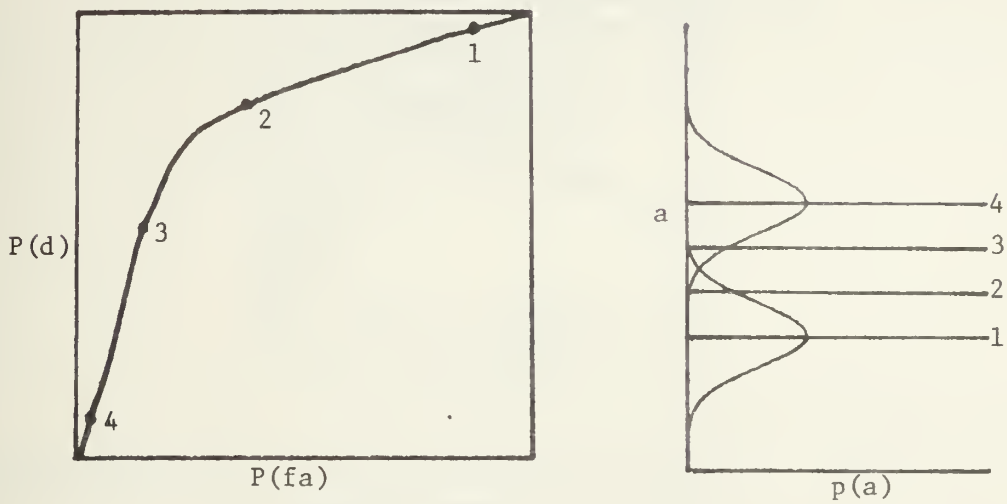


The ratio of P(d) to P(fa) does not remain constant, however. Plotting P(d) versus P(fa) as the threshold varies, the following result is obtained. This is known as a "Receiver



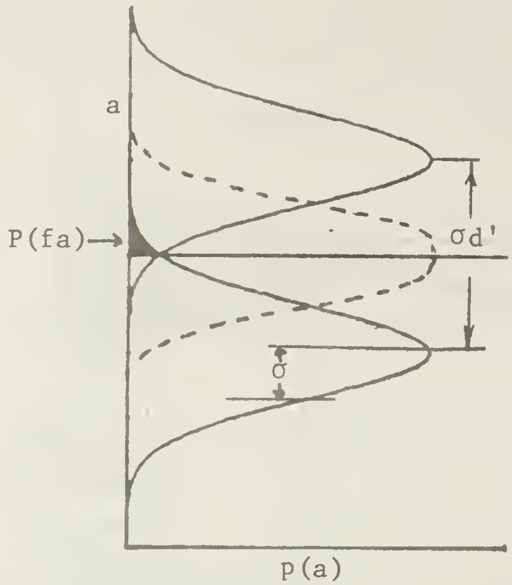
Operating Characteristics" curve or ROC curve. Using this curve it can be determined where to set the receiver threshold to optimize P(d) and P(fa). As the threshold is raised, P(d) does not change much in comparison with the rate at

which P(fa) is decreasing, until the "knee" of the curve (between points 2 and 3) is reached.

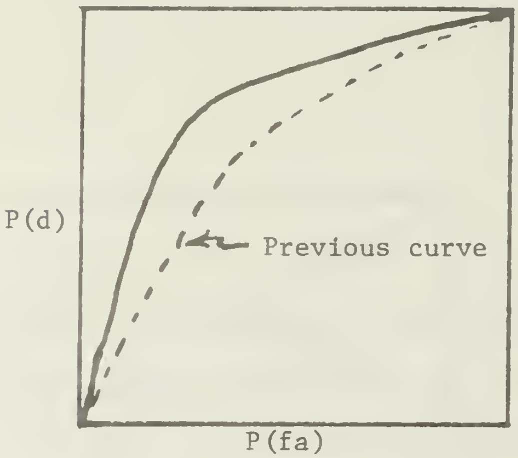


In continuing to raise the threshold above this value, P(d) starts to fall off rapidly, while P(fa) changes very little. By deciding what maximum P(fa) can be tolerated, the threshold can be set to yield the maximum P(d) possible for that particular P(fa) by referring to the ROC curve. Conversely, if a minimum P(d) is required, the ROC curve indicates the minimum P(fa) that must be tolerated.

Until now the effect of changing the mean of the S + N distribution with respect to the N distribution would have on P(d) and P(fa) has not been discussed. If the mean of the S + N distribution is raised, then the ratio P(d)/P(fa) is larger for all threshold settings.

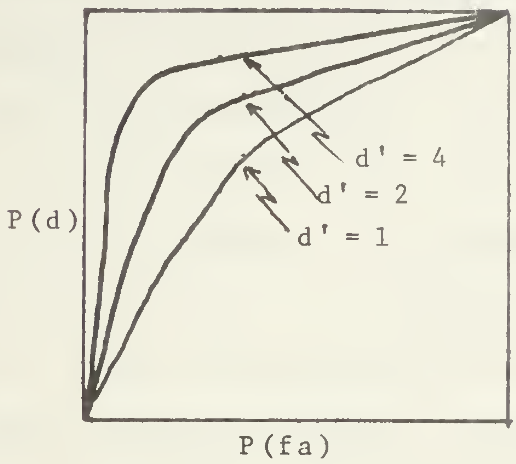


The result is shown on the ROC curve in the drawing below.



"Root detection index" is defined as the quantity d' or $(M_{S+N} - M_n)/\sigma$ which is a measure of the relative separation between the distributions. A large d' value indicates a large separation between the distributions and thus a curve with more of a "knee". The diagram on the following page shows that if the presence of a signal drastically changes the input distribution mean, then the P(fa) can be reduced significantly while still retaining a large P(d). If the

presence of a signal does not drastically change the mean, large values for P(d) are possible only at the expense of a large P(fa).



Thresholds have been discussed and the way the selection of threshold values affects the probability of detection and the probability of false alarm, but how these thresholds are set has not been discussed. One point that may not be readily apparent is that in detection systems in which the human operator is involved in the decision as to presence or absence of signal, his perception of the display must be included in computation of the threshold value. Most systems in use operationally involve an operator in the decision process. Active sonar is an example. The operator can vary the threshold to an extent by adjusting the brightness and intensity of the display and the gain control, but a big consideration in determining whether or not a target is present is his ability to distinguish between target and noise as presented on the PPI scope.

The operator reading spectrograms is another example. He can control the gain, or the grey value of the background on

the paper, but his perception of the distinction between target and noise determines the threshold. For this reason, it is difficult to determine the ROC curves for systems with human operators because it is difficult to determine the actual threshold setting. The threshold will vary from day to day, or even hour to hour, with the same operator and the same hardware settings because of the operator's state of mind, physical condition, fatigue, etc. This is what makes quantification of the detection characteristics for a system difficult. Completely automatic systems are much easier to evaluate, as it can be determined exactly what the settings are.

In the preceding discussion of energy detection and thresholding it has been shown how a system's detection capability can be determined from knowledge of the characteristic distributions of noise and signal plus noise. Implicit in this treatment is the assumption that the distributions of noise and signal plus noise are exactly known. If these distributions are exactly known, the performance of the system can be specified for any given threshold setting in terms of P(d) and P(fa). If these distributions are <u>not</u> known exactly, the analysis will not describe the system performance exactly. Thus, the accuracy of the prediction of system performance depends on the accuracy with which these distributions can be described. In designing real world systems, this is one of the problems facing the designer, since the distributions of noise and signal plus noise are not constant in time (over any lengthy period) or with geographical position. For this reason, many systems

do not always perform as well as the specifications indicate. Only if the ambient conditions are the same as those for which the system was designed will it perform as predicted.

SELF TEST VII

ENERGY DETECTION

Given the distribution functions of noise alone and signal plus noise, with a threshold applied:

1. Probability of detection is given by ______________________.

2. Probability of false alarm is given by ____________________.

3. An ROC curve is a graph of ________________________________

as ____________________________ is varied.

4. Sketch ROC curves for two different signal/noise ratios.

5. It is difficult to include the human operator in an ROC

curve because ____.

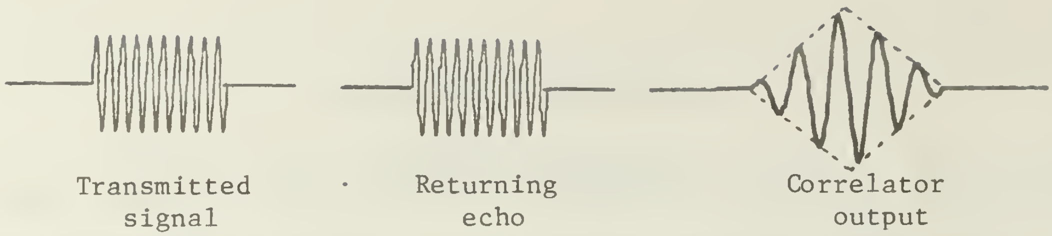
VIII. CORRELATION DETECTION

In Section V, Random Signals, Power Spectral Density, and Noise, it was shown that correlation can effectively improve the signal-to-noise ratio for correlated signals due to the cross product terms in the calculation of the correlation function. The question is how this information can be used to detect signals in noise. As an example, an active echo ranging system is described.

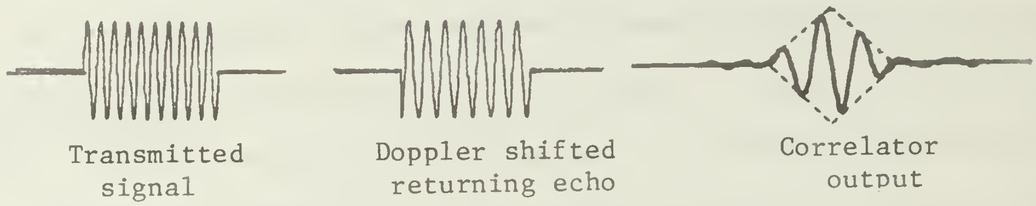
In this type of system a signal is transmitted and then the returning signals are processed. The range of the reflecting object can be determined from the echo return time. In a noise-free environment, there are few problems in this system, but in most real situations the ambient noise may "bury" some of the returns. By using correlation, the signal-to-noise ratio can be improved and these echoes recovered.

This system retains a replica of the transmitted signal and then inputs the received signal into the correlator. The output of the correlator is large when an echo is received and small when only noise is present.

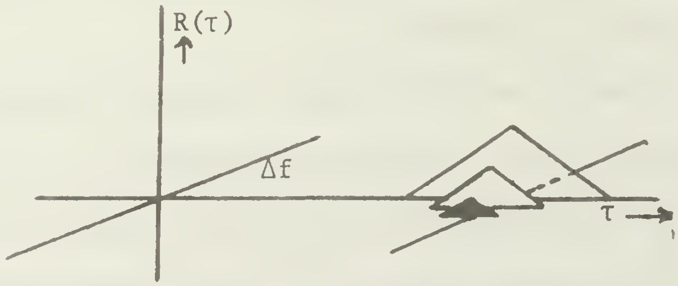
In examining a burst of CW signal transmitted and the effect of doppler shifting on the correlation function, it is noted that if the echo is not doppler shifted the output of the correlator will be the autocorrelation of the transmitted signal at a time τ, which is related to the range of the reflected object. As the returning echo is shifted in frequency by doppler,



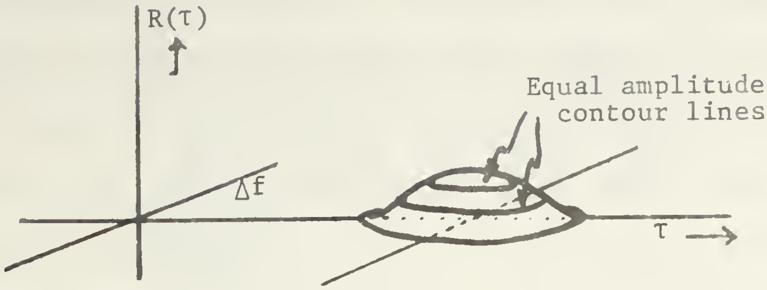
the output of the correlator becomes the cross-correlation of two closely related signals but its value is never as high as the autocorrelation. The cross-correlation amplitude and the amount of delay over which it is correlated decrease. (In fact, the reduction results from the decrease in the delay over which the signal is correlated with its doppler shifted echo).



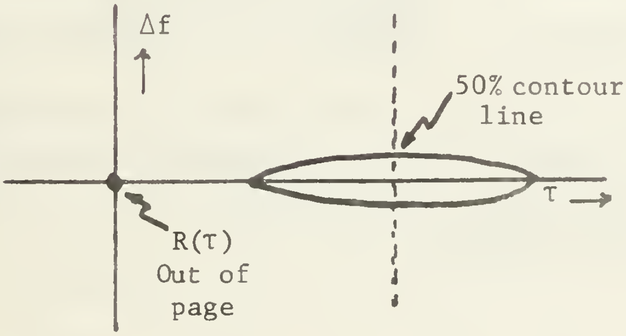
A plot of the correlator output for a number of different doppler shifts on the same diagram, that is a τ, Δf correlator amplitude plot shows the effects of doppler.



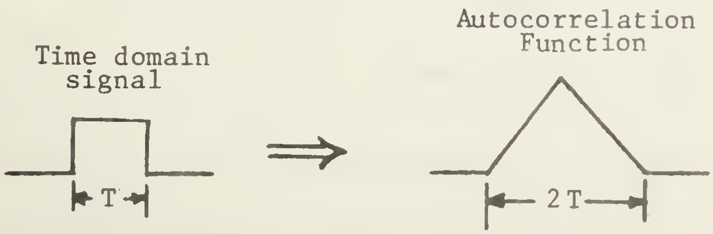
For a given echo, this plot gives a solid surface relating Δf, τ, and amplitude, as shown on the following page.



If at the τ, Δf-plane, the contour line is plotted where the amplitude of the function is 50% of the maximum, the result as shown below is obtained. This diagram is commonly known as an "ambiguity diagram". The reason for this will become apparent from examination of FM pulses. First, consider a "long" CW pulse to see how doppler affects the correlator output.

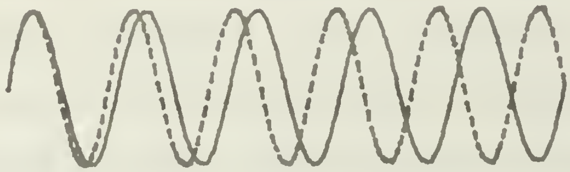


The section on correlation and convolution showed that the width of the correlation function is related to the duration of the time domain signal. Thus for a long CW pulse,

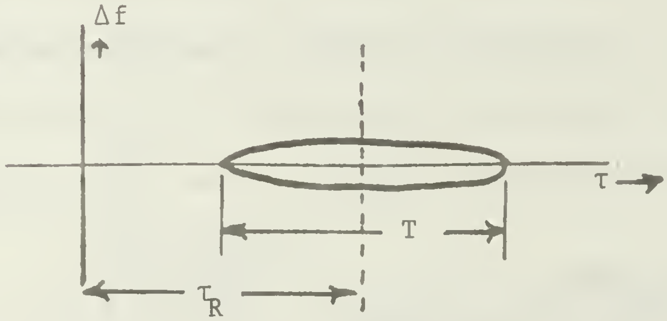


the correlator can be expected to respond over a long range of values of τ.

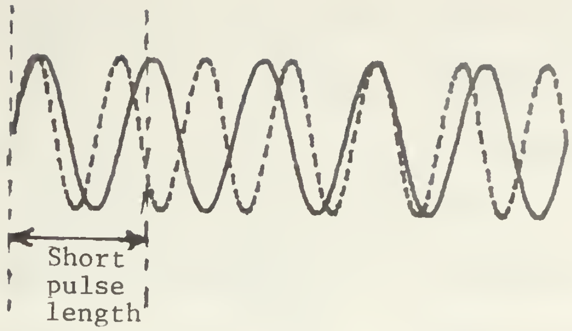
Observation of two signals at slightly different frequencies shows that the longer the pulses to be correlated, the less the correlation, since the signals will be almost correlated over only a few cycles but get further and further out of phase as time progresses. As the difference in frequency between the signals increases, the number of cycles over which the two are almost correlated decreases rapidly. Thus for long



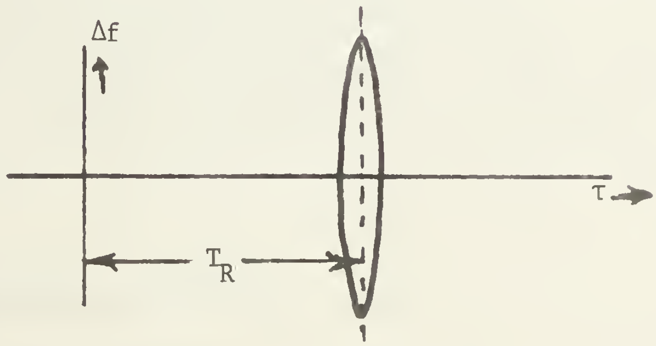
CW pulses, the correlator response decreases rapidly with doppler shifts. The ambiguity diagram for a long CW pulse is shown in the following figure. The length T is proportional to the length of the CW pulse, τ_R is the delay corresponding to the actual range of the target.



For a "short" CW pulse, the length of the correlation peak 50% contour will be shorter, but the range of doppler shifts over which the signals remain essentially correlated is greater.



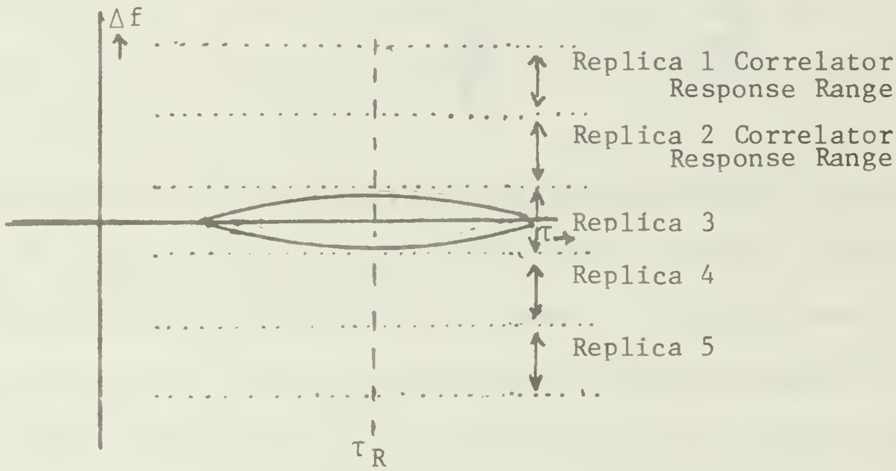
Thus the ambiguity diagram for the short CW pulse has the following appearance:



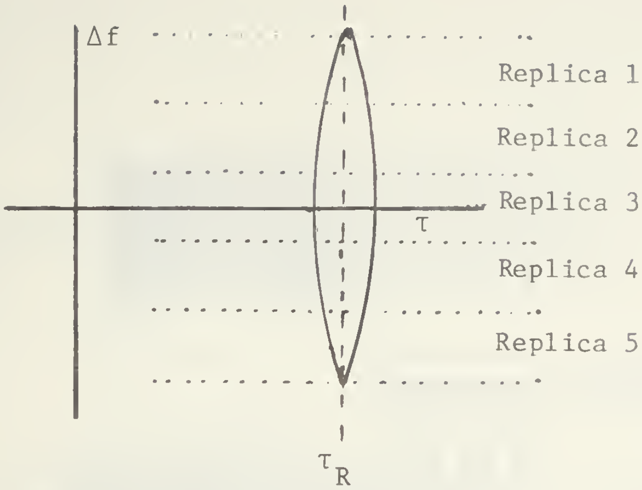
These two diagrams show why long pulse CW is good for searching and doppler determination and short pulse CW is good for range resolution. A sonar pulse of 120 milliseconds will ensonify a region of water 180 meters long. In other words, the spatial length of the pulse as it travels through the water is 180 meters (1500 meters/seconds velocity x 0.120 seconds = 180 meters, the distance the forward edge of the pulse has traveled during the time the rest of the pulse was being generated). The correlation peak 50% contour, as shown on the ambiguity diagram, will be proportional to the pulse length T, and for CW this turns out to be a one-to-one ratio. Thus the correlation will be over a period of 120 milliseconds which means that the target range can be determined only to the nearest 180 meters.

In the case of a short pulse of, for example, five milliseconds, the pulse spatial length is 7.5 meters, the correlation peaks over 5 milliseconds, and the range resolution is 7.5 meters.

The effect of doppler on the ambiguity diagram shows that the same type of development can be made to relate frequency resolution of the two different pulse lengths. Because the long pulse remains correlated over very limited doppler shifts, the frequency of the returning echo can be determined fairly accurately by using a bank of correlators comparing the echo to replicas with varying amounts of doppler shifts. On the

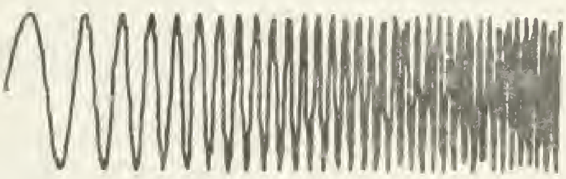


other hand, use of the same bank of correlators with a short pulse produces a response from <u>each</u> of these correlators because the short pulse remains correlated over a large range of doppler shifts, as indicated on the following page.

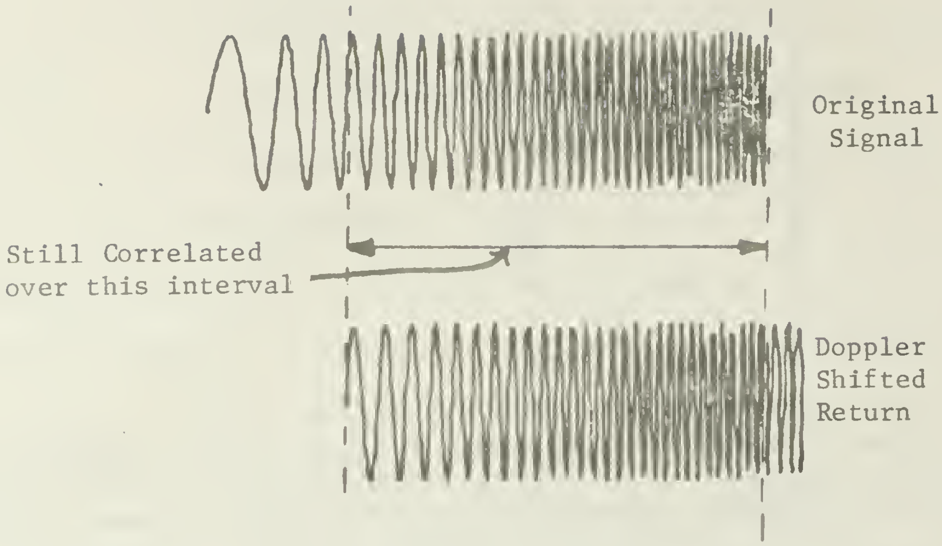


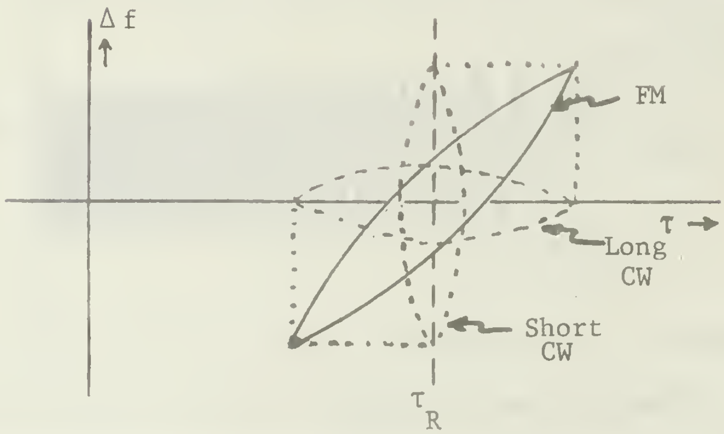
If a pulse is used in which the frequency is changed during transmission, commonly referred to as FM slide, several differences appear. Because of the frequency change during the pulse, the returning echo correlates with the transmitted replica over a short range of time τ, even for long pulses. Thus the

FM "Up-Slide"

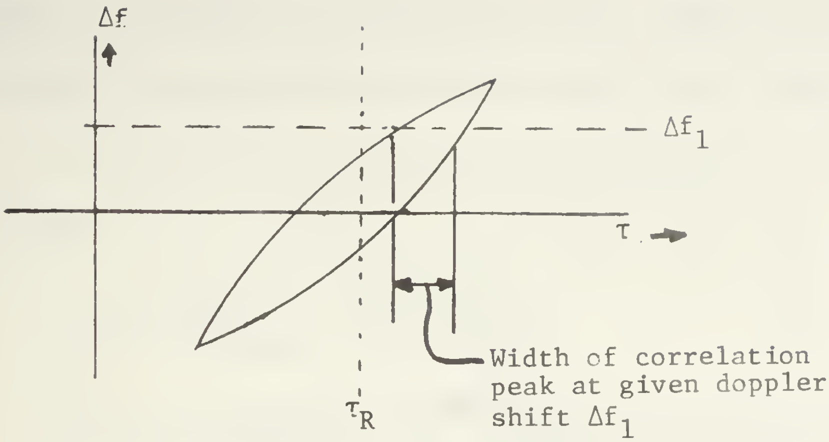


ambiguity diagram is very narrow in the τ direction. On the other hand, if a long pulse is used, the Δf range of correlation is narrow. One difference between FM and CW pulses becomes apparent, however. In the case of FM pulses, a doppler shift will tend to shift the entire pulse in frequency, but the slide will remain linear (provided it was originally linear) and thus it will still tend to correlate but at a different value of τ. This behavior and its effect on the ambiguity diagram are shown in the drawings on the following page.

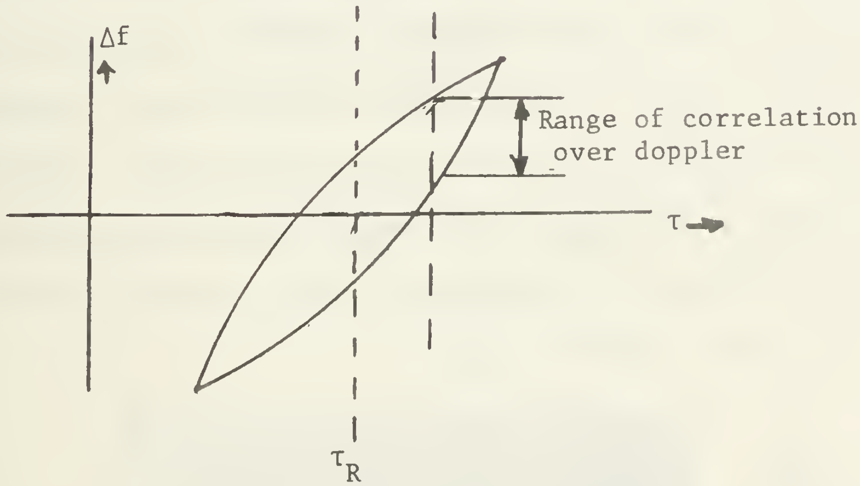




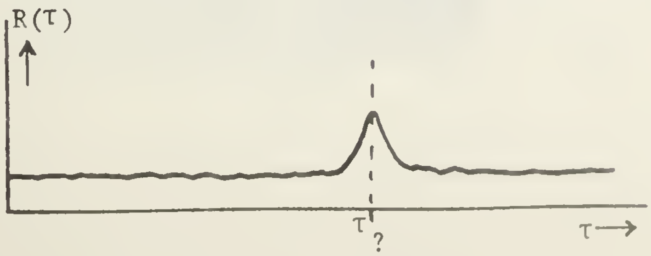
Note that the FM slide pulse has some of the characteristics of both long and short pulse CW. At any given doppler shift, the correlation peak itself is narrow, as is the peak for short CW pulses. Thus, range resolution is good.



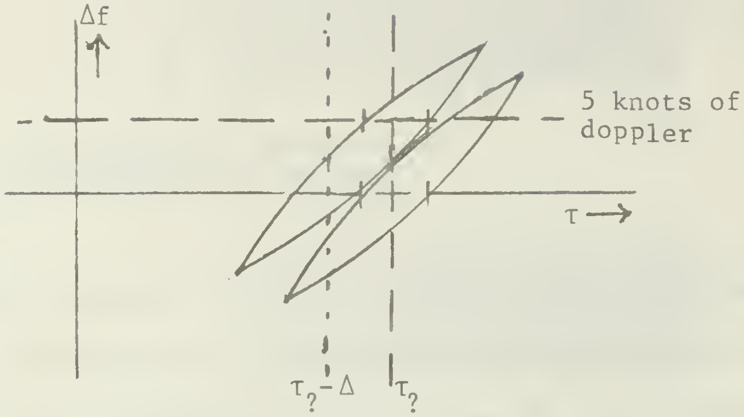
Note also that at a given τ the range of doppler shift over which the signal is correlated is narrow, on the order of the range over which the long CW pulse is correlated. The reason this diagram is called an "ambiguity diagram" can now be seen.



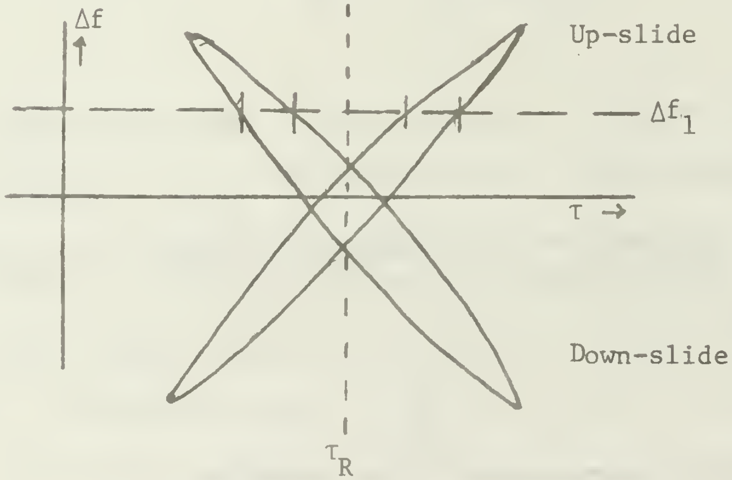
The correlator output, that is, the τ-amplitude plot, is shown in the diagram below. Was the peak at $\tau_?$ caused by a target exhibiting no doppler at range equivalent to $\tau_?$ or was



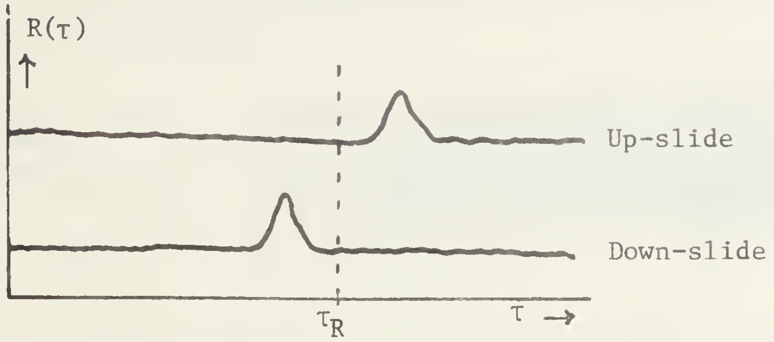
it caused by a target with five knots of doppler at range $\tau_? - \Delta$? The output looks the same in either case. There is a "range ambiguity" present.



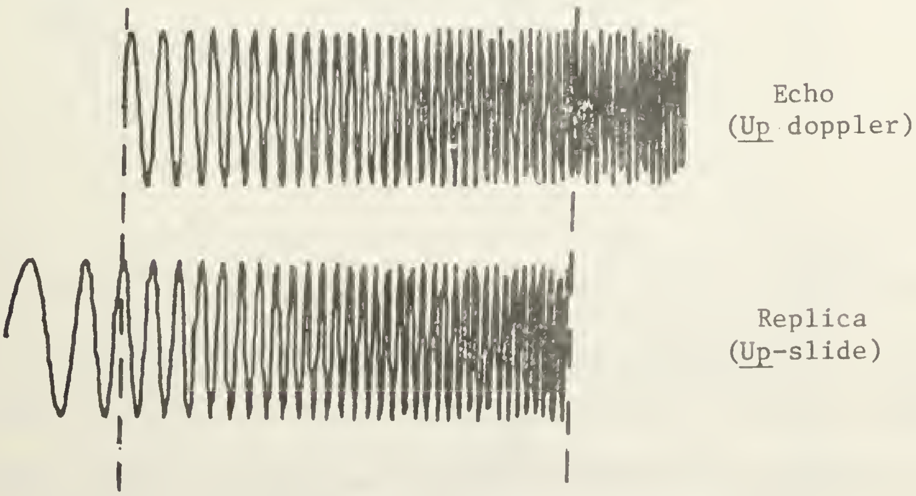
One way to resolve this is to transmit two FM slide pulses, one with an "up-slide" (increasing frequency with time) and one with a "down-slide" (decreasing frequency with time). By then using two correlators, one with an "up-slide" replica and one with a "down-slide" replica, the ambiguity can be resolved. The following ambiguity diagram shows the output from both correlators for a given range target.



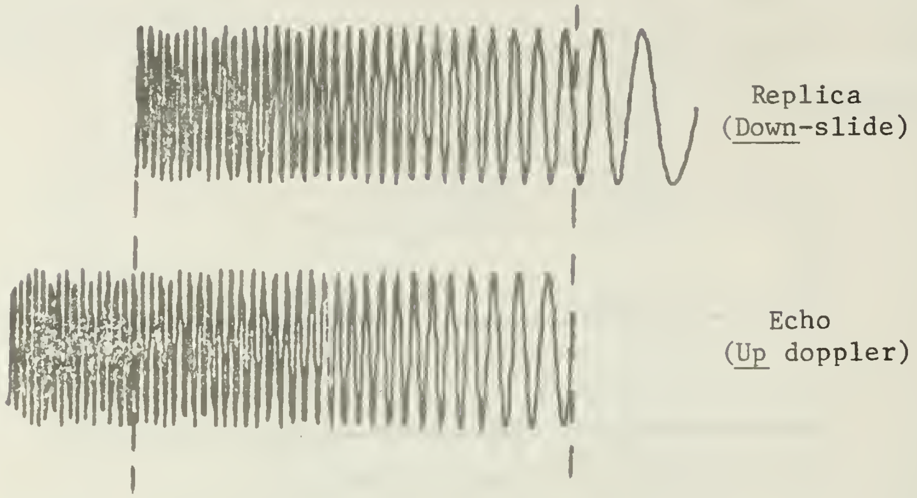
When compared on a τ-amplitude plot, the correlator outputs for a target exhibiting the doppler shift, shown on the ambiguity diagram by the dotted line, appear as in the following diagram.



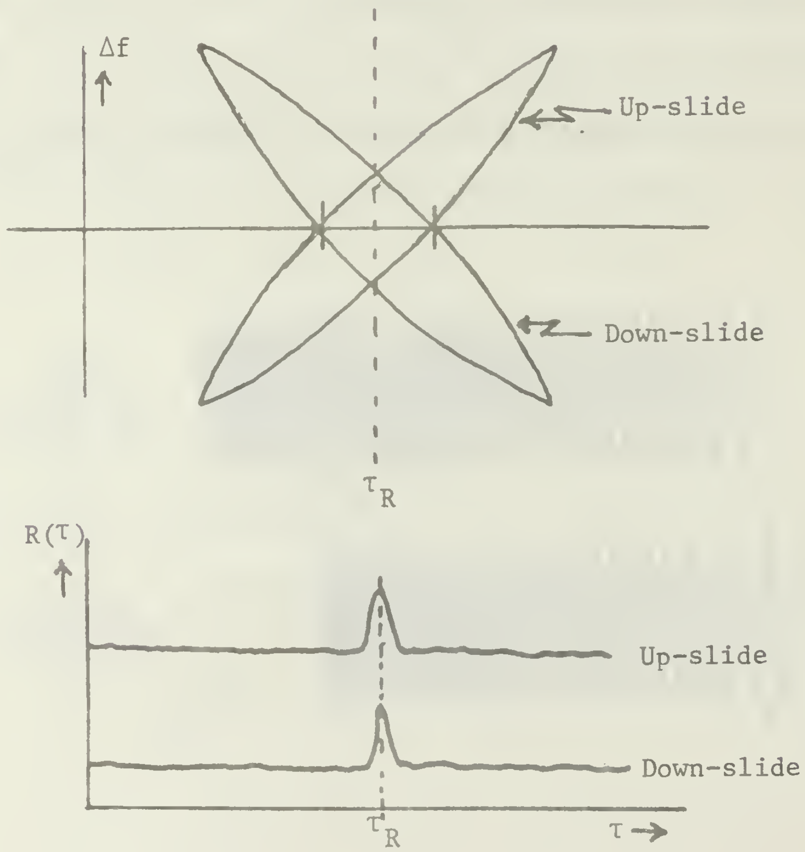
By noting which peak appears first, it can be determined whether the doppler is up or down. If the "up-slide" peak appears last, then the returning echo has been shifted up in frequency and thus correlated at a delay Δ later than it would if not doppler shifted. This indicates up doppler.



If the "down-slide" peak appears first, this confirms the condition which indicates up doppler. The true range of the

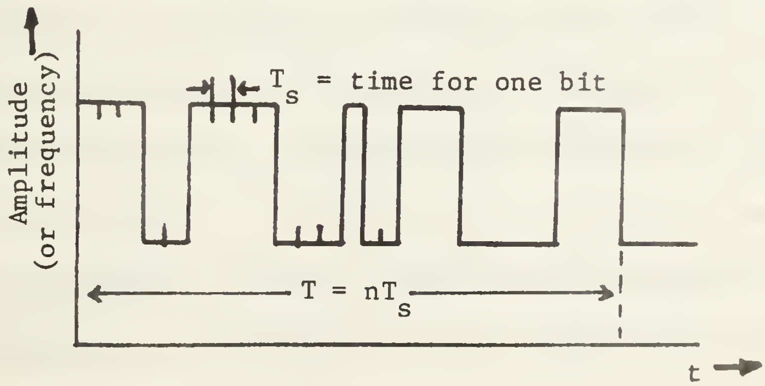


target is midway between the two correlation peaks regardless of the amount of doppler shift present. If the target has no doppler, the two peaks coincide.

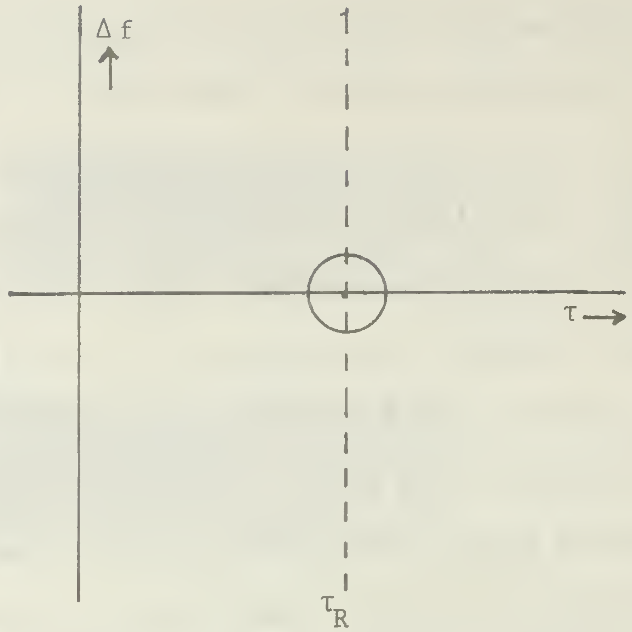


The separation between the up-slide and down-slide correlation peaks is proportional to the amount of doppler shift in the returning echo. If a suitably calibrated display of the two outputs is available, the amount of doppler may be read directly from the presentation. Thus, by employing FM slide pulses and suitable processing, the range resolution of short CW pulses and the doppler discrimination of long CW pulses can be obtained from a pair of long FM pulses.

Another type of pulse which provides the "best of both worlds" with respect to doppler and range is known as the "pseudo-random noise pulse" which consists of a "coded" pulse with amplitude or frequency varied in a pseudo-random fashion, for example, the case where the amplitude (or frequency) of the signal is not a simple sinusoid but a binary coded function.



This pulse is correlated over a very narrow range of τ, giving good range resolution and it is also correlated over a very narrow range of doppler shift because of its pseudo-random character (truly random signals correlate only at a single value of τ). The ambiguity diagram for a pseudo-random pulse is shown in the figure on the following page. Note that because of the

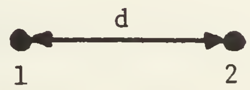


correlation over narrow ranges of f and τ, no range ambiguity is present. However, the pseudo-random pulse requires a separate correlator for each doppler channel which is more expensive than dual FM. Also, another problem associated with this type of pulse is that the amplitude of partial correlation peaks in R(τ) may be greater than for FM and tends to degrade the false alarm probability and thus degrades the performance of real systems.

Up to this point the use of correlation in active systems and how the use of FM slide pulses combined with the proper correlation processor makes it possible to determine the amount of doppler in an echo directly have been covered. Examination of some applications of correlation in the passive detection filed will be discussed.

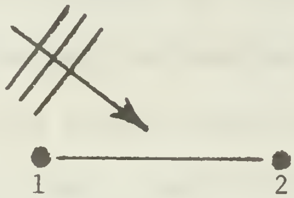
One difference between active and passive correlation detection systems becomes apparent very quickly. That is, in

the active system the returning echo is compared with a replica of the transmitted pulse. In other words, the signal sought is a known quantity (a replica of the transmitted pulse). In a passive system, however, no pulse is transmitted, therefore there is no replica for comparison with the incoming signal. Consequently, correlation cannot be utilized in the same manner; however it is still of use, particularly in passive direction-finding systems. The utility of passive direction-finding systems can be seen by examination of a typical system and its operation. Consider a system composed of two hydrophones separated in space by some distance d. If the output

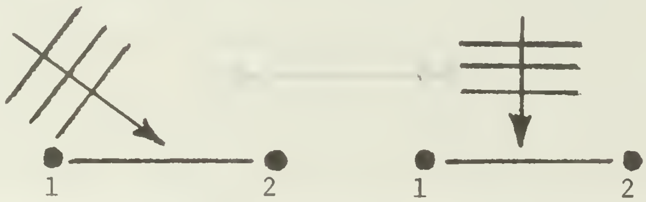


of hydrophone 1 is used as a replica with which to compare the output of hydrophone 2, the cross-correlation of the two outputs can be obtained. The important property of sea noise for this system is that it tends to be isotropic, that is, to be the same regardless of the direction from which it impinges on the hydrophone. In addition, sea noise tends to be uncorrelated over distance. On the other hand, most signals of interest are from a single source, usually a submarine, and thus are coherent to an extent and directional in nature. Since sea noise is isotropic, the correlator output will be small and relatively constant when no signal is present. In the presence of a signal of interest, the acoustic energy traveling outward from the

source will impinge on the hydrophones from some given direction. Following a given wavefront as it travels across the hydrophones, it is seen that it will encounter one of the



hydrophones before the other unless the direction of impingement is perpendicular to the line joining the two hydrophones or the baseline. The difference between the times that the

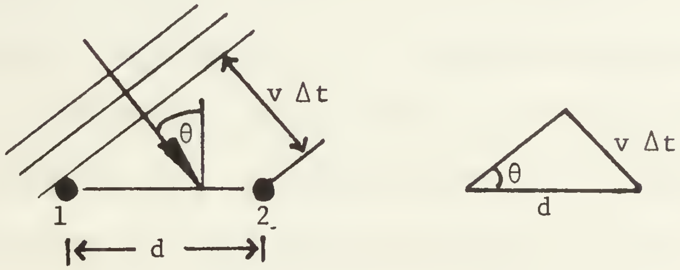


wavefront strikes the two hydrophones depends on the direction from which it arrives with respect to the baseline. If the direction of arrival is perpendicular to this baseline, the wavefront strikes both hydrophones at the same time and thus there is no time difference. If the direction of arrival is parallel to the baseline, the time difference is a maximum. In fact, the time difference is related to the arrival direction by the following equation:

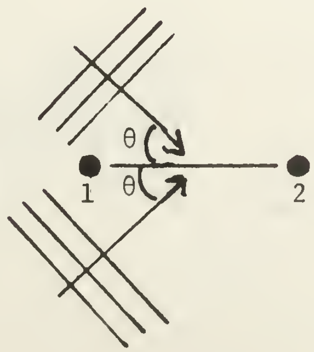
$$\frac{v \Delta t}{d} = \sin \theta$$

where v is the speed of sound in the medium and θ the wavefront

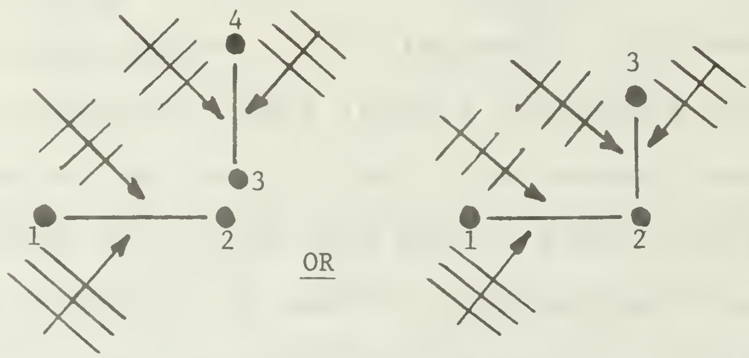
arrival angle with respect to broadside. The cross-correlation of the two hydrophone outputs for a single wave passing the array shows that the delay time τ at which a correlation peak appears is a function of the wavefront arrival time difference at the hydrophones. If the wavefront arrival angle

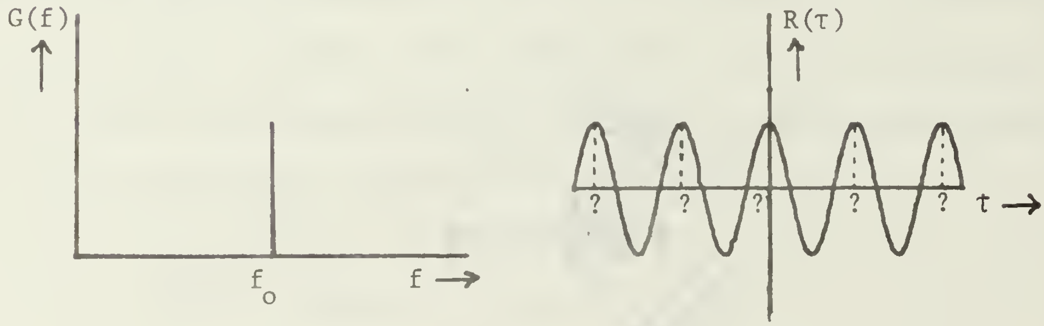


is 0°, then the arrival time difference is zero and thus the correlation peak will appear at $\tau = 0$. For any other arrival angle, there is a positive arrival time difference Δt and the correlation peak appears at $\tau = \Delta t$. This allows relating the correlation delay time τ to the wave arrival direction with respect to the array baseline. There is a problem of directional ambiguity as arrivals from opposite sides of the array at equal angles with respect to the baseline produce the same delay.



This ambiguity can be resolved by adding another array to cross-fix, or by adding a third hydrophone not in line with the first two. It should be noted that although this method works well with broadband radiated signals, it has some drawbacks when applied to signals of a single frequency line or those with a very broad autocorrelation function. The problem is that if a signal consists of a single frequency line, its autocorrelation function is a sinusoidal function. The question then arises as to which peak correponds to the signal arrival direction. For this reason, a correlation detector is not well suited for passive detection of narrow-band signals, but works well for broadband signal detection.





It has been seen how correlation can be of help in active systems and in passive broadband signal detection systems. What can be done about single frequency line detection? In order to answer this question, a system will be examined which is similar to correlation in many aspects but is not actually correlation.

SELF TEST VIII

CORRELATION DETECTION

1. A correlator processes the received signal by __________.

2. Sketch the output of a correlator for a CW pulse signal.

3. Doppler causes what effect on the output of a correlator for a CW pulse signal?

4. What is an "ambiguity diagram"?

5. Compare the ambiguity diagrams for short pulse and long pulse CW signals.

6. Doppler causes what effect on the output of a correlator for an FM pulse signal?

7. What are the advantages of an FM pulse over a CW pulse?

8. How can range-doppler ambiguity be resolved in an FM processor?

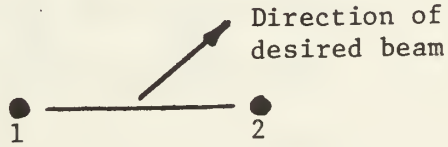
9. What are the disadvantages of a pseudo-random pulse?

10. How can a correlator be used as a passive direction finder?

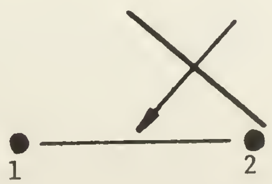
11. Why is a passive correlator's performance degraded for narrow band signals?

IX. BEAM FORMING

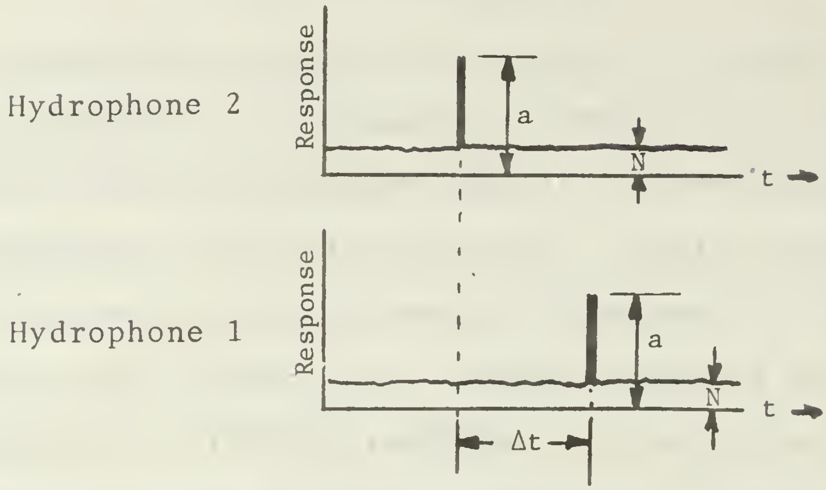
Beam forming, or electrical steering of arrays, is accomplished by a process which is similar to correlation but with one major difference -- it uses summation instead of multiplication of the outputs. Starting with a two-hydrophone array, the process is examined to understand how it operates. As noted in the preceding section, the acoustic wave arrival direction determines the wavefront arrival time difference at the hydrophones. For example, assuming that it is desired to "steer" the array to receive signals from a given direction relative to the array baseline selectively, that is, to form a "receiving beam" at a given orientation, if the output is delayed from Hydrophone 2 for a time $t_d = \Delta t$, and add t_d to the time necessary for the acoustic wave from the desired direction to travel from Hydrophone 2 to Hydrophone 1, the array can be selectively biases to receive signals from that direction.



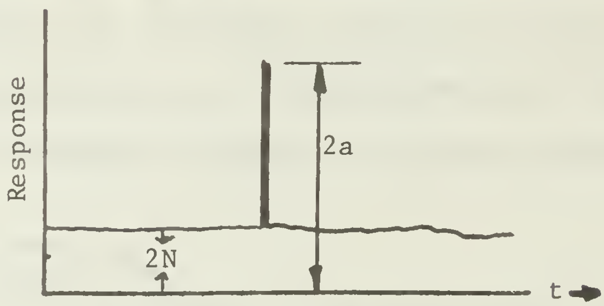
Why does this selectively bias the response of the system to signals from that direction? Consider a single impulse arriving from the direction of the selected beam.



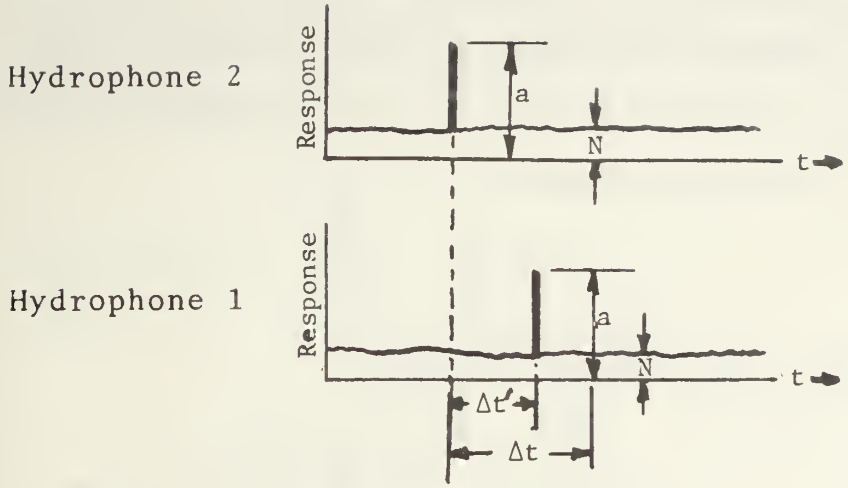
Plotting the output of each hydrophone as a function of time produces the following result:



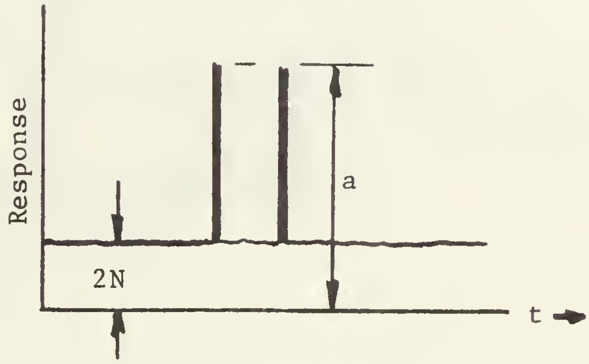
By delaying the output of Hydrophone 2 for a time $t_d = \Delta t$ and adding the output of the two hydrophones, the two impulses reinforce each other and the output will appear as in the following:



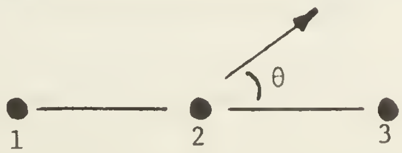
If, however, the impulse arrived from a different direction, the time difference between arrival at the two hydrophones will be different as shown in the illustration on the following page.



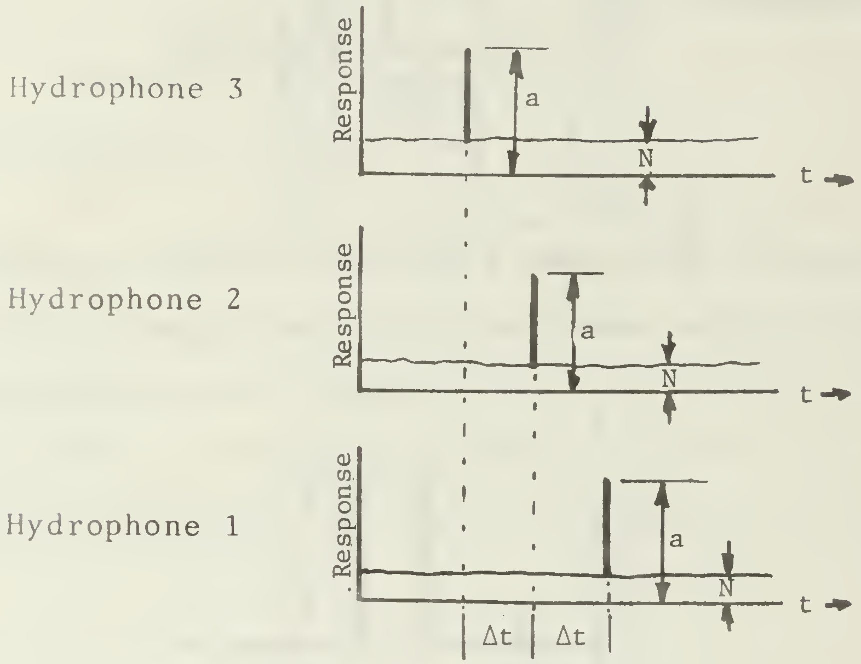
If the output of Hydrophone 2 is input through the same delay Δt, the plot of the summation of the two signals is as shown below:

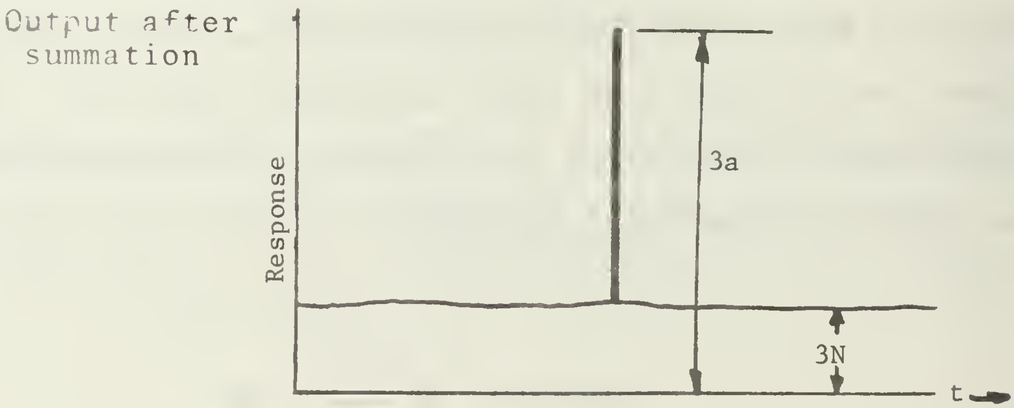


Thus the signals arriving from the selected direction (time delay Δt) will selectively reinforce one another in the processor, while those from other directions will not. If a 3-hydrophone linear array is examined, it is observed that the effect is even more pronounced.



By delaying the output of Hydrophone 3 for a length of time $2\Delta t$ and the output of Hydrophone 2 for a length of time Δt, a wave from the desired direction will produce the effect shown in the following figures:



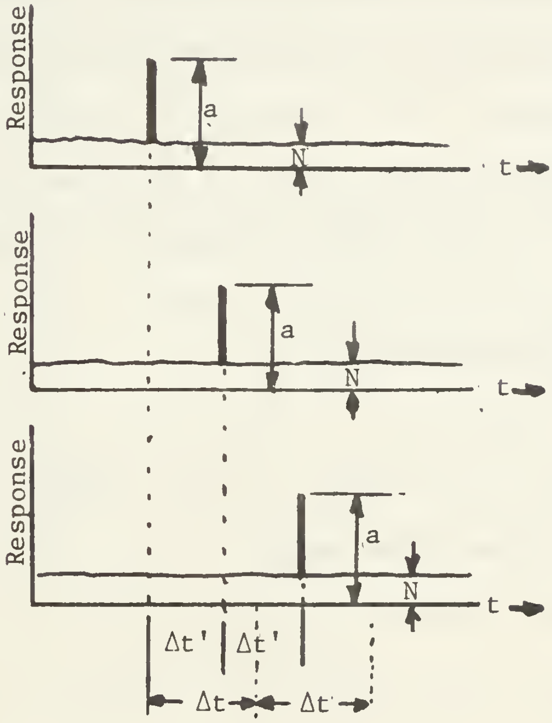


A wave from a different direction will give the following output.

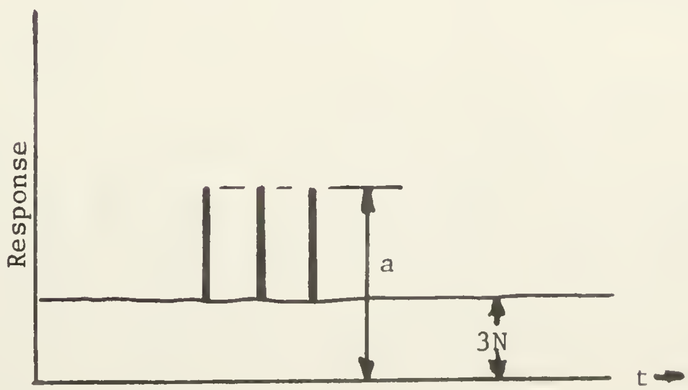
Hydrophone 3

Hydrophone 2

Hydrophone 1



Output after summation



It is apparent that adding hydrophones to the array makes it more responsive to signals from the chosen direction. Also, since the output of the system is not a correlation of the individual hydrophone outputs but a simple summation of their amplitudes, the system will work for narrow-band or single-frequency line signals as well as for broadband signals.

The signal power from the desired direction is given by the square of the entire summation of amplitudes:

$$S = \left[a_1 + a_2 + a_3 \right]^2$$

$$= a_1^2 + a_2^2 + a_3^2 + 2a_1 a_2 + 2a_1 a_3 + 2a_2 a_3 = 9a^2$$

since all the a's are equal.

The noise power is similarly:

$$N = n_1^2 + n_2^2 + n_3^2 + 2n_1 n_2 + 2n_1 n_3 + 2n_2 n_3$$

However the noise amplitudes are randomly positive and negative, so that on the average, the crossproducts add up to zero.

$$n_1 n_2 + n_1 n_3 + n_2 n_3 \approx 0$$

Therefore, in this case $N \approx 3n^2$ and the signal/noise ratio in the desired direction is 3 times that of an omni hydrophone, or in general:

$$\left(\frac{S}{N}\right)_{\text{array}} = \left[\begin{array}{c}\text{number of elements} \\ \text{in the array}\end{array}\right] \times \left(\frac{S}{N}\right)_{\text{omniphone}}$$

By using a number of hydrophones in this manner, in conjunction with an energy detector, and by setting the threshold high enough to preclude response to the unreinforced signals from other directions, one has a very directional detection capability from a series of omnidirectional hydrophones.

WAVE DIFFRACTION - FOURIER THEORY

Transmitted beam patterns may be analyzed as the phenomenon of diffraction. The concept of diffraction is based on the addition of phasors. Waves may have the mathematical form of a phasor function of distance r and time t. The waves from any pair of phase-locked (coherent) sources will add vectorially depending on the phase difference.

$$\text{Phasor Wave} \qquad A\,e^{j2\pi f\left(t-\frac{r}{c}\right)}$$

where A = amplitude

$2\pi f\left(t-\frac{r}{c}\right)$ = phase

f = frequency of wave

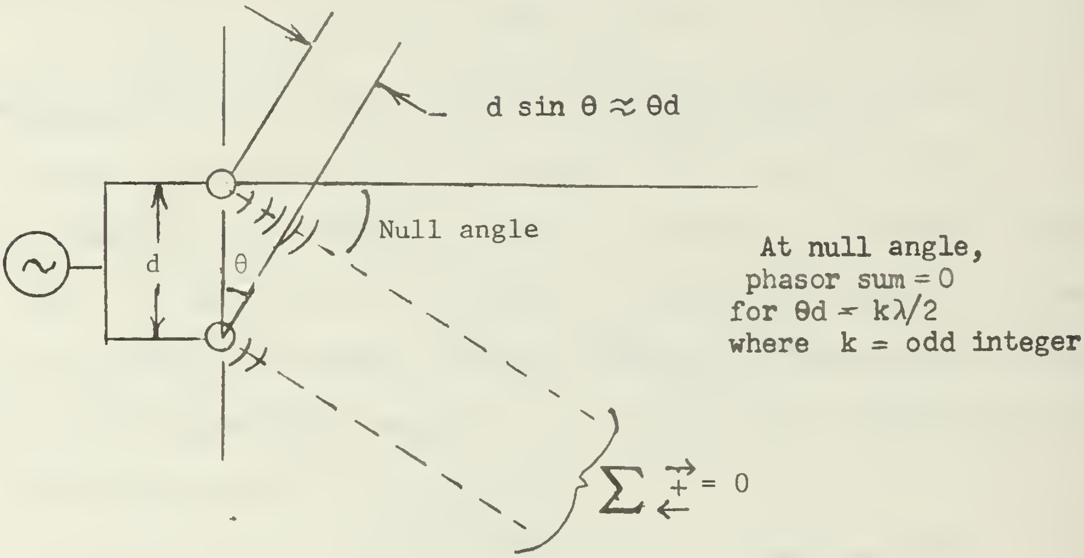
c = velocity of wave

NOTE: $f\lambda = c$

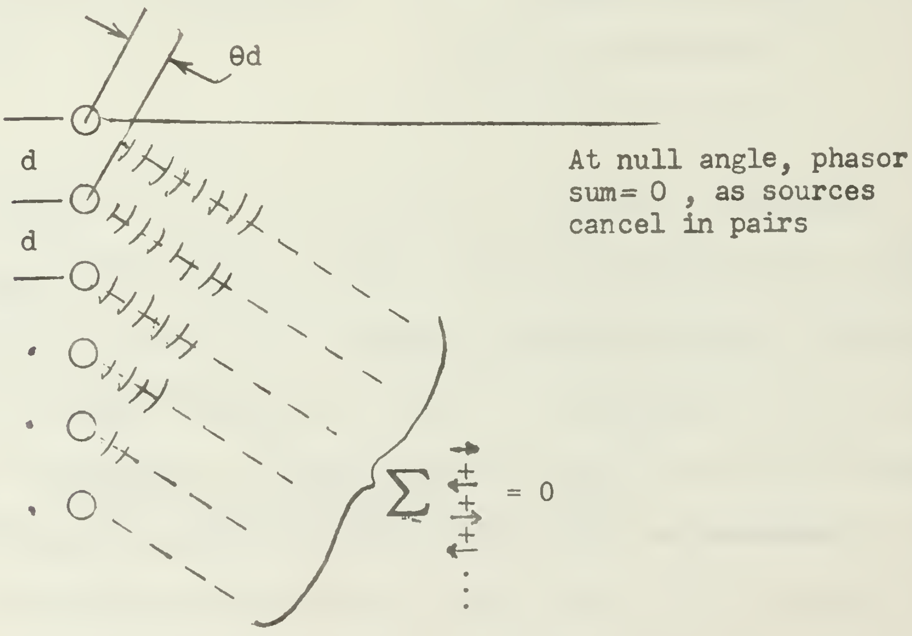
λ = wavelength

A classical example of diffraction is the dipole or "double slit" experiment in which a sinusoidal signal excites two point sources spaced a distance d apart. The waves observed at a distance r from the dipole where r is large compared to both d and the wavelength λ are found to interfere constructively at certain angles and destructively at other angles. The null angles exist for conditions in which waves from the two sources travel a distance in which the

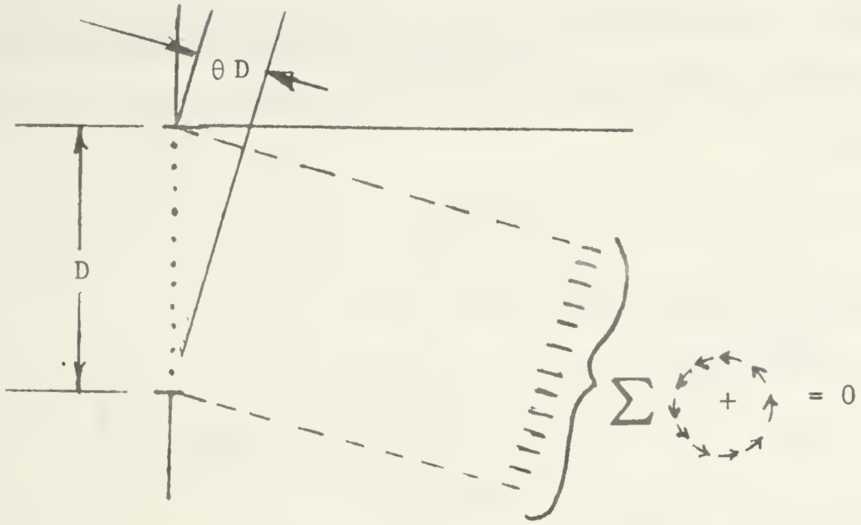
contribution from each source is out of phase with the other. The phasor sum is therefore zero at the null angle. This corresponds to the condition sketched in which the wave from one source is delayed by a distance $d \cdot \sin\theta$ (or approximately θd for angles less than 20°



Another example is the "grating" (many sources equally spaced).



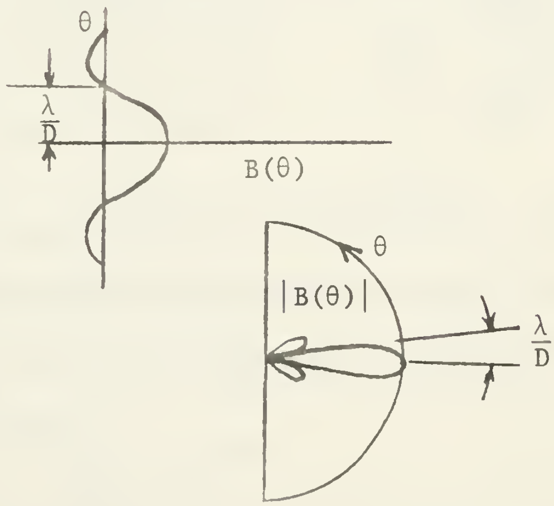
A third example is the "clear aperture" (a single source of width comparable to a few λ).



The total radiation leaving an aperture at a given angle is, therefore, the summation (integral) of all contributions, delayed in phase an appropriate amount. For the clear aperture,

$$B(\theta) = \int_{-D/2}^{+D/2} [1]\, e^{-j2\pi \frac{\theta x}{\lambda}}\, dx$$

$$= D[1]\, \frac{\sin\left(\frac{2\pi}{\lambda}\frac{D}{2}\theta\right)}{\left(\frac{2\pi}{\lambda}\frac{D}{2}\theta\right)}$$



Therefore, any other arbitrary aperture function A(x) produces a radiation pattern of the form,

$$B(\theta) = \int_{\text{Aperture}} A(x)\, e^{-j\frac{2\pi}{\lambda}\theta x}\, dx \qquad \text{where } A(x) = \text{aperture function}$$

The last integral equation can now be recognized as the FOURIER TRANSFORM of A(x), where $S \approx \frac{\theta}{\lambda}$ (actually, $S = \frac{\sin\theta}{\lambda}$). Therefore, an arbitrary aperture function can be said to have a SPACIAL FREQUENCY SPECTRUM proportional to the RADIATION PATTERN.

$$\boxed{\begin{array}{l} B(s) = \displaystyle\int_{\text{Aperture}} A(x)\, e^{-j2\pi sx}\, dx \\ \\ \text{or} \\ \\ B(s) \longleftrightarrow A(x) \end{array}}$$

where $\boxed{\theta = \lambda s}$

Now consider a "sinusoidal grating" in which the aperture is excited by a distribution of amplitude given by the equation,

$$A(x) = A_o \cos(2\pi\, x/x_o)$$

The radiation will be diffracted into only two beams at an angle $\pm\theta_o$ from the normal to the grating. The angle θ_o is given by the formula

$$\sin\theta_o = \frac{\lambda}{x_o} \quad \text{or} \quad \theta_o \approx \frac{\lambda}{x_o} \quad \text{for small angles.}$$

Therefore, the diffracted radiation angle is inversely proportional to the spacial period x_o, or directly proportional to the spacial frequency.

$$\boxed{\text{SPACIAL FREQUENCY } s_o = 1/x_o}$$

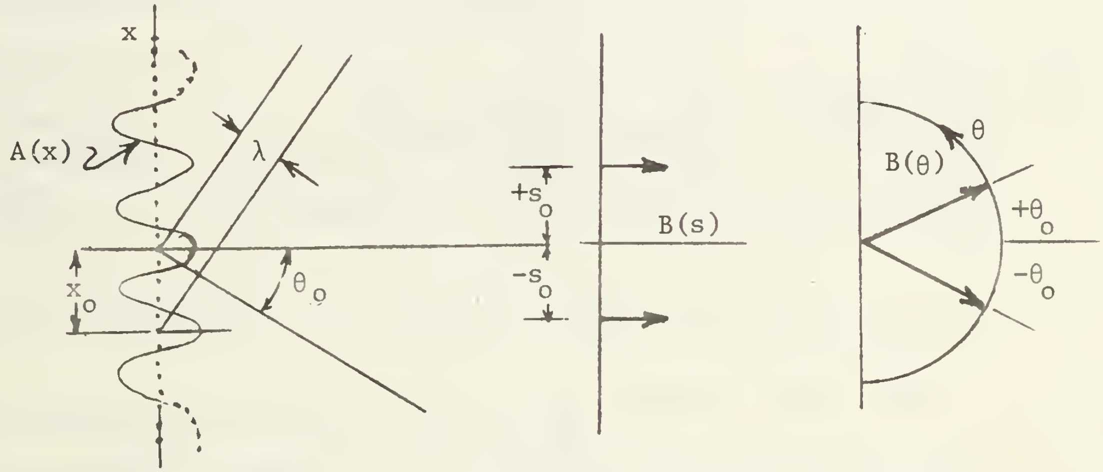
then

$$A(x) = A_o \cos 2\pi s_o x$$

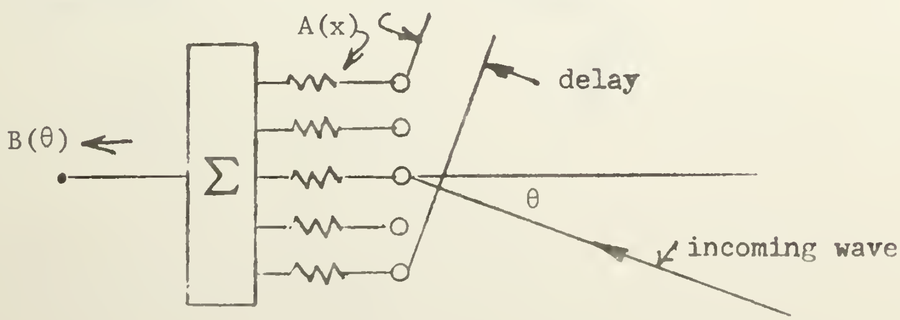
and the diffraction angle is given by

$$\theta_o = \lambda s_o$$

being linearly proportional to the spacial frequency s_o. Therefore there is a one-to-one correspondence between the spacial frequency at the aperture and the angle of diffraction. The sinusoidal grating has only a single spacial frequency s_o. Any aperture other than a sinusoidal grating has a continuous distribution of spacial frequencies.

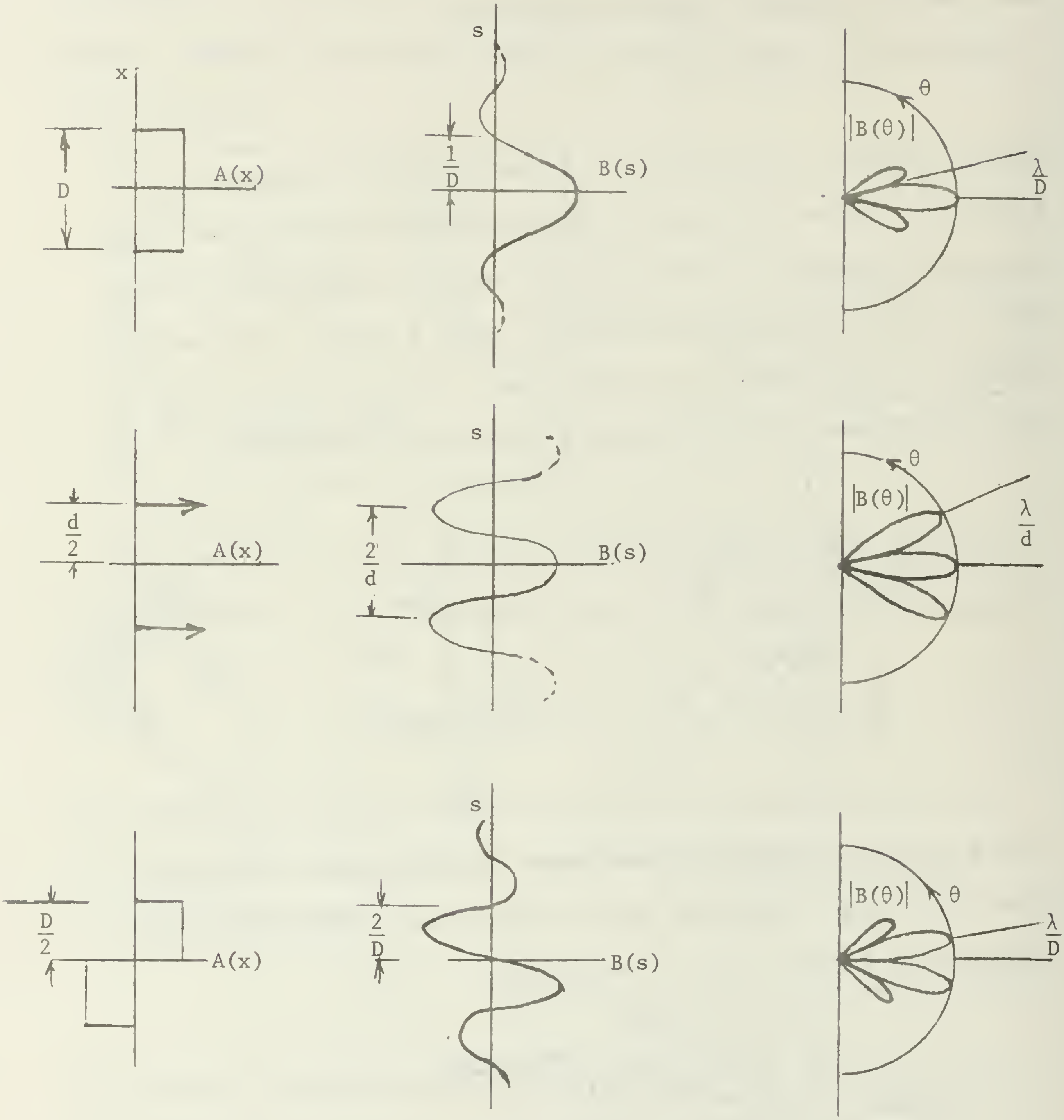


Note that a receiving beamformer accomplishes the same phase shifting (delaying) and integrating (summing) as the diffraction process in transmitting.



Therefore, the same mathematics applies and the receiving pattern is the same as the transmitting pattern.

EXAMPLES:



SELF TEST IX

BEAM FORMING

1. Beam forming and electrical steering of a receiving array is accomplished by ______________ and ____________ the output of each element.

2. The signal/noise ratio at the output of a beam former can be approximately equal to a factor __________________ better than the S/N of a single element.

3. Transmitted beam patterns may be analyzed as the phenomenon of ____________________.

4. The transmitted beam pattern is related to the excitation function of the aperture by _______________________.

5. Spacial frequency is _______________________________.

6. The receiving beam pattern is the same as the transmitting pattern because ________________________________.

BIBLIOGRAPHY

Bracewell, R., The Fourier Transform and Its Applications, McGraw-Hill, 1965.

Carlson, A. B., Communication Systems, McGraw-Hill, 1968.

Decker, T. R. and Sollenberger, R. T., The Environment and Its Effect on Antisubmarine Warfare, Masters Thesis, Naval Postgraduate School, Monterey, 1975.

King, L. F. Jr. and Swift, D. A., Acoustic Primer for Underwater Sound Applications to Antisubmarine Warfare, Masters Thesis, Naval Postgraduate School, Monterey, 1975.

Sakrison, D. J., Communication Theory: Transmission of Waveforms and Digital Information, Wiley, 1968.

Smith, B. J. and Vice, L. W., Probability and Its Applications to Antisubmarine Warfare, Masters Thesis, Naval Postgraduate School, Monterey, 1975.

Taub, H. and Schilling, D. L., Principles of Communication Systems, McGraw-Hill, 1971.

Urick, R. J., Principles of Underwater Sound for Engineers, McGraw-Hill, 1967.

INITIAL DISTRIBUTION LIST

	Copies
Dean of Research Code 012 Naval Postgraduate School Monterey, California 93940	1
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
Commanding Officer Air Test and Evaluation Squadron 1 (VX-1) Patuxent River, Maryland 20670	1
Commanding Officer Code 7000 Naval Torpedo Station Keyport, Washington 98345	1
Commanding Officer Submarine Development Group Two Groton, Connecticut 06340	1
Commanding Officer Naval Air Development Center Warminster, Pennsylvania 18974	1
Center for Naval Analysis 1401 Wilson Boulevard Arlington, Virginia 22209	1
Commanding Officer Naval Weapons Center China Lake, California 93555	1
Director Naval Weapons Laboratory Dahlgren, Virginia 22448	1
Commanding Officer Naval Surface Weapons Center Silver Spring, Maryland 20910	1

Director 1
Naval Research Laboratory
Washington, D. C. 20390

Commander 1
Naval Ship Research and Development Center
Bethesda, Maryland 20034

Commander 1
Naval Ocean Systems Center
San Diego, California 92152

Commanding Officer 1
Naval Intelligence Support Center
4301 Suitland Road
Washington, D. C. 20390

Commander 1
Naval Electronics Systems Command
2511 Jefferson Davis Highway
Arlington, Virginia 20360

Commanding Officer 1
Attn: Underwater Sound Laboratory
Naval Underwater Systems Center
New London, Connecticut 06320

Commander 1
Naval Ship Engineering Center
Hyattsville, Maryland 20782

Commanding Officer 1
Naval Coastal Systems Laboratory
Panama City, Florida 32401

Commander 1
Naval Air Systems Command, Code 370
Washington, D. C. 20361

Commander 1
Naval Sea Systems Command, Code 03424
Washington, D. C. 20362

Commanding Officer 1
Naval Underwater Systems Center
Newport, Rhode Island 02840

Commanding Officer 1
Naval Ordnance Station
Indian Head, Maryland 20640

Commanding Officer Naval Surface Weapons Center Dahlgren, Virginia 22448	1
Commander Anti-Submarine Warfare Systems Projects Office Washington, D. C. 20360	1
Director Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217	1
Naval Postgraduate School Code 62Sa Monterey, California 93940	150

NPS